Jagaputhran S

Senior ML Engineer / Architect

📞 +91 8807224113

✉ jagaputhran.s@gmail.com

📍 Chennai, TN

Experienced Machine Learning Engineer with over 7.5 years of extensive experience in developing, deploying, and optimizing machine learning models and data engineering pipelines. Proficient in Databricks, Snowflake, Python, Spark, and SQL, with a strong background in data engineering and visualization using Tableau dashboards. Adept at client-facing roles with excellent communication skills, ensuring effective collaboration and delivery of tailored solutions. Experienced in both traditional ML and generative AI/LLM projects and driving product-oriented visions. Demonstrated expertise in managing and optimizing data workflows, building robust data models, and implementing scalable data solutions on cloud platforms.

WORK EXPERIENCE

Senior Machine Learning Engineer

Comcast | April 2024 - Present

- Developed "*Decoding Anomalies in Big Data: Unmasking Obfuscated Command Line Threats in CrowdStrike Using a Custom BERT and Unsupervised Models*"
- Submission of one **IDD** (under review) and actively working on 2 ideas
- Actively working with **MLflow** and related tools for an in-house developed syslog parser
- In-house project that focuses on *building tools that simplify or automate core pipeline and platform engineering tasks using LLMs in Databricks*
- Leveraging the capabilities of in-house Databricks models, utilizing **deep learning and neural networks** for model understanding
- CIEC Lab week ideas: **Adsense AI** and **Sec QR**, which received high praise from senior leadership
- Built **chatbots** for streamlining processes and a proprietary model for cyber teams
- **GenAI ideation challenge** submissions
- <u>Mentorship:</u>Training 2 interns in data-science/statistics and 2 interns, 1 FTE in data engineering from scratch, providing guidance and support to help them develop their skills and knowledge in these fields

Development Engineer 3

Comcast | July 2021 - March 2024

- Worked predominantly as a **full-stack Pyspark Developer** on Databricks, Snowflake, PySpark, and **CI/CD** pipelines using **Concourse**
- Orchestrated jobs using Apache Airflow with a deep understanding of AWS
- As part of the **Cybersecurity team (TITAN)**, worked on flattening various security feeds and logs and contributed to various in-house development projects
- Worked on the initial phases of **Databee**
- Developed numerous **ETL** pipelines to streamline data processing and integration
- Worked with **SKY** team for **Delta Live tables** integration

Associate

Cognizant | Nov 2017 - July 2021

- Delivered complex projects using Agile and Scrum methodologies.
- Optimized business logic using Spark-SQL and DataFrames; scheduled pipelines with Apache Airflow, YARN, Cloudera, AWS EMR
- Basic knowledge of Kafka, Spark Structured Streaming, and NoSQL databases (Apache Cassandra, MongoDB).
- Proficient in handling JSON, CSV, and Text file formats.
- Conducted business and data analysis with Power BI; knowledgeable in retail and e-commerce processes.
- Developed a linear regression-based classifier to predict outcomes in the retail client projects

EDUCATIONAL BACKGROUND

Master's(M.Tech) degree in Data Science and Engineering
Birla Institute of Technology and Science
2022 - 2024

Post Graduate Diploma in Product Management(PGDPM)
Institute of Marketing Communications India
2021 - 2022

Bachelor's degree in Electronics and Communication Engineering
St.Joseph's College of Engineering
2013 - 2017

TECHNICAL SKILLS

Machine Learning Algorithms: Linear Regression, Logistic Regression, Decision Trees, Random Forests, XGBoost, SVM, k-NN

Deep Learning: Neural Networks (CNN, RNN, LSTM, Transformers), GANs, Autoencoders

Natural Language Processing (NLP): Text Preprocessing, Tokenization, Word Embeddings, Named Entity Recognition, Sentiment Analysis

Computer Vision: Image Processing, Object Detection

Statistical Analysis: Hypothesis Testing, Probability, Bayesian Methods

Data Preprocessing: Feature Engineering, Dimensionality Reduction, Data Normalization

Deployment: Model Optimization, A/B Testing, CI/CD for ML Pipelines

Cloud Platforms: AWS, GCP (specific to ML services like SageMaker, Vertex AI)

TOOLS/SOFTWARE

Programming Languages: Python (NumPy, Pandas, Scikit-learn), R, SQL

Deep Learning Frameworks: TensorFlow, PyTorch, Keras

Data Visualization: Matplotlib, Seaborn, Plotly, Tableau, Power BI

Big Data Tools: Apache Spark, Hadoop

Version Control: Git, GitHub, GitLab

Containerization and Orchestration: Docker, Kubernetes

CI/CD Tools: Jenkins, GitHub Actions, MLflow

Databases: MySQL, PostgreSQL, MongoDB, Snowflake

IDEs: Jupyter Notebook, PyCharm, VS Code

NLP Libraries: Hugging Face, SpaCy, NLTK

Model Deployment: Flask, Streamlit, Gradio , Chainlit ,Django

Cloud Platforms: AWS SageMaker, GCP Vertex AI, Mosaic AI

Experiment Tracking: MLflow, Weights & Biases

CERTIFICATIONS

AWS Certified Machine Learning Speciality - Amazon Web Services

CompTIA Security+ - The Computing Technology Industry Association (CompTIA)

Oracle Cloud Infrastructure 2024 Generative AI Professional Certification - ORACLE