ARAVIND BABU SOMEPALLI

ARTIFICIAL INTELLIGENCE / MACHINE LEARNING ENGINEER
somepalliaravindbabu@gmail.com | +1 806-445-9562 | Lubbock, TX | LinkedIn| GitHub

Summary

Machine Learning Engineer with over 3 years of experience in Artificial Intelligence, Deep Learning, Machine Learning, and Data Mining. Skilled in developing and deploying large-scale, production-ready applications, particularly in the domain of Natural Language Processing (NLP) and advanced machine learning algorithms. Adept in building models utilizing Large Language Models (LLMs) and deploying those using platforms like Databricks. Demonstrated expertise in Python for data manipulation, extraction, and model training, using libraries such as NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Ggplot2, OpenCV, PyTorch, and NLTK. Experienced in managing a broad suite of AWS cloud services, including EC2, S3, Lambda, EBS, EMR, DynamoDB, SQS, SNS, and CloudWatch, ensuring scalable, secure, and cost-effective machine learning operations in cloud environments.

Professional Experience

United Health Group, USA | MACHINE LEARNING ENGINEER | Jan 2025 – Present

- Designed and implemented a scalable chatbot framework using DAG architecture and LangChain agents with Voyager + ReACT prompting, improving conversation flow accuracy and reducing support workload by 30%.
- Enhanced dialogue automation for ReACT chatbot products, resulting in an estimated cost savings of $300K annually.
- Fine-tuned Large Language Models (LLMs) using PEFT and LoRA techniques to automate customer support workflows, leading to an 8% improvement in KPIs such as resolution time and CSAT.
- Leveraged Hugging Face Transformers (BERT, GPT, ALBERT, CLIP, BART) to develop and optimize NLP pipelines for text classification, summarization, and sentiment analysis tasks.
- Deployed LLM applications across AWS (SageMaker, EC2, Lambda, EMR, S3) and GCP environments, achieving 99.95% system uptime with seamless cross-platform scalability.
- Improved observability of ML workflows by integrating AWS CloudWatch and alerting systems, reducing issue resolution time by 15% and enhancing uptime for critical models by 25%.
- Conducted A/B testing and causal inference experiments to evaluate new ML-powered web features, leading to a 12% increase in user engagement.
- Upgraded deep learning pipelines by refining CNN and Transformer architectures, which boosted model performance by 15% in real-time classification use cases.

Gyandata, India | Data Scientist | Jan 2021 – Jul 2023

- Processed and analyzed ~20M rows of procurement data using Apache Spark on Databricks and Snowflake, uncovering actionable insights to optimize daily equipment usage.
- Implemented K-means clustering models to segment equipment based on consumption behavior, reducing maintenance costs by 7–8% across industrial operations.
- Engineered 20+ lag-based features and deployed Gradient Boosted Regression models for time series forecasting, improving procurement planning accuracy by 21%.
- Led the integration of AWS services (EC2, Lambda, S3) with on-premises systems to create a hybrid cloud platform, cutting operational costs by 29% and increasing infrastructure flexibility.
- Managed large-scale relational and non-relational databases including SQL Server, PostgreSQL, MySQL, MongoDB, and NoSQL for structured and unstructured data storage.
- Conducted sentiment analysis on customer emails using LSTM-based RNN models, achieving 89% sentiment classification accuracy and enabling targeted service improvements.
- Optimized TensorFlow models for compute efficiency by implementing pruning and quantization, enhancing model performance by 22% and reducing resource consumption by 26%, enabling deployment of 10+ production-ready models.

Technical Skills

- **Language**: Python, R, Java, SQL, MATLAB
- **IDEs**: Visual Studio Code, Anaconda, JupyterLab, Jupyter Notebook
- **ML Algorithm**: Linear Regression, Logistic Regression, Decision Trees, SVM, Random Forests, Naive Bayes, K Means, Supervised, Unsupervised, Clustering, NLP, Large Language Models (LLM), LSTM, NLTK
- **Deep Learning:** ANN, CNN, RNN, Hugging Face Transformers (BERT, GPT, ALBERT, CLIP, BART), Stable Diffusion
- **Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, PyTorch, NLTK, Ggplot2, OpenCV, Plotly
- **Cloud Technology:** Azure, AWS (SageMaker, EC2, EMR, S3), GCP(Pub/Sub, Big Query, Composer)
- **Visualization Tools:** Tableau, Power BI
- **Web Experimentation:** Causal Inference, A/B Testing, Hypothesis testing, Statistical Modeling, Optimization, Experimental Design, Docker, Flask, CI/CD pipelines
- **Big Data/Database**: SQL Server, MySQL, NoSQL, MongoDB, PostgreSQL, PySpark, SparkSQL, Spark, ETL Data Pipeline, Hadoop, SVN, Snowflake, Databricks, Hive
- **Other Tools & Skills:** GitHub, Jira, Selenium, FastAPI, Airflow, Looker, Excel, PowerPoint,
- **Operating System**: Windows, Linux, MacOS

Education

Master of Science in Computer and Information Sciences|Texas Tech University, Lubbock, TX, USA | Aug 2023 – May 2025
Bachelor of Technology in Computer Science and Engineering|Kalasalingam Academy of Research and Education, Tamil Nadu, India | Jun 2019 – May 2023