Naga Kiran Reddy Karnati

New York | nagakiranreddy2@gmail.com | +17165339305 | GitHub | LinkedIn | Kaggle

Summary

- Data scientist with **5 + years** of expertise in developing and optimizing **Data and Machine Learning** pipelines.
- Extensively worked on **financial** data, including banking, **credit card**, and wealth management advisory datasets, building insightful **dashboards** for leadership to drive strategic decision-making.
- Skilled in utilizing data visualization tools like **SAS, Tableau, Power BI and Looker** for creating reports and dashboards.
- Expert in utilizing various packages such as Scikit-learn, Pandas, SciPy, Seaborn, Matplotlib, NLTK, and spaCy
- Worked on diverse projects involving **Generative AI with Large Language Models**, utilizing PEFT and LoRA for efficient training and enhanced productivity.
- Proficient in deep learning frameworks such as **TensorFlow and PyTorch**, with expertise in implementing RNNs, CNNs, and attention mechanisms for various applications.
- Consistently updated with **Gen-AI** news and **RAG** studies, applying the latest methods for superior model performance.

Experience

TD Bank, Bank of Montreal (Essex Lake Group LLC, New York) **Apr 2023 - Current**

Solution Intelligence Specialist

- Innovatively implemented **50+** controls for **TD Bank** using **SAS and Excel** to identify and prevent regulatory and policy breaches in the **US credit card** business, ensuring compliance and operational integrity.
- Developed and optimized reusable code templates in **SAS and Python**, reducing runtime by **40**% and manual intervention by 30%, while ensuring seamless scalability for future cloud migration to **Azure**.
- Defined risk thresholds and conducted data-driven analysis on **5M+** records, enabling leadership to assess control risks with 99% accuracy and implement corrective actions by collaborating with **clients** on flagged exceptions.
- Mentored and guided team members through knowledge-sharing sessions and reusable code resources, accelerating onboarding and enhancing project delivery **timelines**.
- Led data preparation efforts by transforming raw datasets through **SQL and Excel pivots,** ensuring seamless integration for **BMO's** PB (Private Banking LOB) project dashboards.
- Converted **20+** code runs into a one-click process with built-in **QC** checks and automated email alerts for code run status, enhancing operational efficiency and reliability.
- Implemented advanced **SQL** techniques like macros and sub queries for automating data ingestion and reduced process run-time by **67% from 6 to 2 hours.**

Researchwire Knowledge Solutions Pvt. Ltd, Navi Mumbai, India **Jul 2019 - Aug 2021**

Data Scientist

- Implemented transformer architecture for Named-Entity Recognition (NER) task, boosting F1 score significantly by fine-tuning pre-trained models with tokenizers from **HuggingFace** Library.
- Orchestrated end-to-end **machine learning workflows** by integrating AWS SageMaker with AWS Lambda, S3, and Glue, achieving a **60%** improvement in model training efficiency and saving **$50,000** annually in operational costs.
- Developed question-answering (QA) model using **transformer** networks, achieving a 20% accuracy increase over baselines by fine-tuning pre-trained transformers and implementing algorithms in **TensorFlow** and PyTorch.

NIT Surat, India **Jan 2018 - Jun 2019**

Data Scientist Intern | ADCS Group Head

- Engineered a Character-Level Language Model using Recurrent Neural Networks (RNNs), mitigating the vanishing/exploding gradient problem by **25%** through **gradient clipping** and employing novel sampling for human-like text fluency.
- Optimized sequential data processing with transformer networks, RNNs, GRUs, and LSTMs, enhancing analysis tasks by 30%.
- Employed **Python and R** for advanced analytics and **statistical modeling**, harnessing libraries such as Pandas, NumPy, Matplotlib, and Scikit-learn to derive insights.

Projects

GenAI with LLMs **Jan 2023 - Feb 2023**

Python,, AWS Sagemaker, LLMs

- Utilized **data parallelism and Microsoft's ZeRO optimizations** to manage large-scale models, reducing computational costs by 40% without compromising performance.
- Applied Reinforcement Learning from Human Feedback (**RLHF**), increasing model alignment with human values by **50%**, as evidenced by user satisfaction.
- Deployed Retrieval-Augmented Generation (**RAG**), ensuring more precise answers and reducing **hallucination**.
- Leveraged the **LangChain** framework, boosting productivity and workflow efficiency.

Fleet Scheduling using Multi-Agent DDPG **Apr 2022 - May 2022**

Python, Open AI gym, Multi-Agent Reinforcement Learning

- Implemented Multi-Agent DDPG algorithm for vehicle scheduling, boosting orders served by **20%.**
- Designed a **10000x10000** grid environment simulating real-world scenarios for multiple agents (vehicles) to learn the optimal routes, achieving a 30% reduction in average delivery time.
- Explored off-policy learning algorithms such as **Q-learning and SARSA** to train agents from historical data, enabling efficient policy improvement and value estimation.

Transfer Learning using MobileNet_v1 **Feb 2022 - Mar 2022**

Python, TensorFlow, Scikit-learn, OpenCV

- Employed transfer learning with MobileNet_v1 for image classification tasks, achieving an accuracy of **92%** on a diverse dataset.
- Fine-tuned pre-trained MobileNet_v1 on target domains, reducing training time by **50%** while maintaining high accuracy.
- Implemented model interpretability techniques such as Grad-CAM to visualize the regions of interest, aiding in model debugging and validation.
- Collaborated with domain experts to customize MobileNet_v1 architecture for specific **image** recognition requirements, resulting in tailored solutions for different applications.

Optimizing Deep Neural Networks **Dec 2021 - Jan 2022**

Python, TensorFlow, Scikit-learn

- Experimented with optimization methods such as SGD, Momentum, RMSProp, and Adam, achieving a **15%** improvement in model convergence speed.
- Utilized random minibatches to **accelerate convergence and improve optimization**, resulting in a **25%** reduction in training time.
- Implemented learning rate decay scheduling, doubling the training process speed.
- Explored ensemble learning methods such as **bagging and boosting** to combine multiple models, resulting in a **15%** enhancement in predictive accuracy.

Education

Master of Science in Machine Learning and IoT **Aug 2021 – Feb 2023**

University at Buffalo, SUNY, Buffalo, NY, CGPA: 4/4

B. Tech in Electronics and Communications Engineering **Jul 2015 – May 2019**

NIT Surat (SVNIT), Surat, India, CGPA (Top 4): 9.04/10

Skills

Languages/IDEs: Python, R, Java, C, C++, SQL, Pyspark, Scala, Matlab, HTML, CSS, Jupyter, Google Colab, VS Code

Libraries/Modules: TensorFlow, PyTorch, Scikit-learn, Keras, LightGBM, Pandas, NumPy, Matplotlib, Seaborn, Plotly

NLP: Gen AI, LLMs, RNN, Attention Mechanisms, Sentiment Analysis, Speech Recognition, NLTK, spaCy, Hugging Face

Big Data Tools: Big Query**,** Apache Hadoop, Apache Spark, Kafka, Airflow, Hive, AWS/GCP/Azure (Cloud Platforms), VetexAI

Others: Tableau, Power BI, Microsoft Excel, Git, SAS, MySQL, SQL Server, PostgreSQL, Docker, Kubernetes, Databricks, Leadership

Certifications

- **Machine Learning** Specialization
- **Deep Learning** Specialization
- **Generative AI** with LLMs
- **AWS ML**
- **Natural Language Processing (NLP)** Specialization
- Advanced **SAS SQL,** Power BI, **AWS DS** Essentials