# Nikhil Nageshwar Inturi

+1 (945)216-4026 | inturinikhilnageshwar@gmail.com | https://www.linkedin.com/in/nikhilinturi/



**SUMMARY:**

Data Scientist and Machine Learning Engineer with 7+ years of experience in data-sriven applications and scalable solutions. Proficient in Python, GenAI, PyTorch, ETL pipelines, cloud(AWS, Azure), containerization (Docker, Kubernetes), and CI/CD workflows, with a strong track record in optimizing workflows and collaborating across teams. Passionate about solving complex problems with AI and Data Science.


**SKILLS:**

**Programming:** Python(NumPy, Pandas, SciPy), R, SQL, Shell Scripting, Java, Workflow languages (Cromwell and NextFlow), DQL
**Machine Learning and AI:** Deep Learning(Keras and PyTorch), NLP(RNN, LSTM, Transformers), Hugging Face, Generative AI(LlamaIndex, LangChain)
**Containerization and Deployment CI/CD:** Git, Docker, Podman, Kubernetes, Jenkins, GitLab CI, and Docker Swarm
**Databases and Cloud Tools:** Redis, PostgreSQL, Snowflake, Redshift, Azure Cosmos DB, AWS DynamoDB, MongoDB, and **AWS** and **Azure** (*certified*)
**Certifications:** Databricks GenAI Fundamentals, Graduate Certificate in Applied Machine Learning, Post Graduate Program in AI and Machine Learning, Amazon Cloud Computing Practitioner, Microsoft Certified: Azure AI/Data Fundamentals

**EXPERIENCE:**

**Senior Data Scientist, The University of Texas at Dallas** Feb 2023 – Present

- Developed classification models using **XGBoost** & **LightGBM** for rat jaw-size prediction, improving predictive performance in VNS simulation experiments by 87%.
- Led a team of 5 in building image segmentation models (**Detectron2**, **yoloV11**) increasing the neuron detection by 15% and reducing processing time by 95%.
- Developed **clustering model**s, processing 70M+ reads across ~4900 features, using **shared nearest neighbor** clustering.
- Developed a **GPT-powered RAG pipeline** (LangChain) to enable semantic search across research papers and datasets(**S3**, **MariaDB**) at CAPS, integrating database queries for associated analyses, reducing research retrieval time by 70%.
- Collaborated with researchers at McGill (Canada) and University of Queensland (Australia) to streamline **NGS workflows**.

**Data Scientist, Aganitha Cognitive Solutions** Jun 2022 – Nov 2022

- Developed **clustering models(K-means, GMM)** to identify AAV capsid sequence identification that cross the blood-brain-barrier, resulting in a 99.96% reduction in required in-vivo experiments.
- Built an **interactive analysis tool** to process 30M+ records for AAV Capsid Engineering, integrating **Python**, **Cromwell**, **Bash**, **R**, and **RESTful APIs** for data visualization workflows.
- Fine-tuned a deep learning model (Splice-AI - spliceai5) to improve novel splice junction detection in humans, leading to a ~30% reduction in false positives and improving downstream biomarker identification.
- Enhanced genome search efficiency by developing **advanced search algorithms** for AutoBLAST, achieving 2x faster performance compared to the traditional BLAST search.

**Data Scientist, Infosys Ltd** Sep 2018 – Jun 2022

- Built **40+ database connectors** (SQL, NoSQL, Cloud) in Python, improving data accessibility for ML pipelines.
- Automated data ingestion & processing pipelines using Pandas, PySpark, and Airflow, reducing data preparation time by 30% while ensuring data quality, traceability, and 100% code coverage with robust testing.
- Integrated classification and clustering algorithms (Scikit-learn, LightGBM, CatBoost, H2O, AutoML, Keras) into Infosys Data Science Platform (IDSMLP), enhancing model automation and predictive capabilities of the tool.
- Automated SAP CFIN (S/4 HANA) reporting in Python, eliminating manual data extraction and reducing report generation time by 40%.
- Implemented Containerization strategies using docker for AutoML platform to streamline deployments.

**Data Science Project**, PURDUE University Nov 2020 – Oct 2021

- The project enhances stock market prediction using Deep Reinforcement Learning (DRL), Natural Language Processing (NLP), and LSTM to analyze historical data and sentiment from news/social media.
- The model achieves 96.8% accuracy in sentiment analysis, improves stock price prediction by 80%, and outperforms traditional strategies with a Sharpe ratio of 3.0 and ARR of 1.1.

**EDUCATION:**

**The University of Texas at Dallas,** *Master's in Business Analytics & Artificial Intelligence* | GPA: 3.9 Dec 2024
**Purdue Global – Simplilearn,** *Post Graduate Program in AI and Machine Learning* | GPA: 10 Nov 2021
**Ramaiah Institute of Technology,** *Bachelor of Engineering in Mechanical Engineering* | GPA: 9.8 June 2018