# Naveen Morla

(419) 333 1085 | naveen.morla04@gmail.com | linkedin.com/in/Naveen-Morla | github.com/naveenmorla1901

## WORK EXPERIENCE

**Data Scientist / Software Engineer** Jan 2024 – Present

*THK Manufacturing of America, Inc.*

- Engineered high-performance APIs using **FastAPI** with advanced caching for seamless data ingestion, processing, and retrieval of machine and inventory datasets, and developed and fine-tuned predictive models with **PyTorch** and **TensorFlow** to enhance lead time forecasting and optimizing inventory management.
- Executed comprehensive data preprocessing, feature engineering, and data wrangling to improve model accuracy and robustness, leveraged **Polars** and **PySpark** for accelerated data processing, ensuring efficient handling of large datasets.
- Architected and managed containerized **microservices** with **Kubernetes** and **Docker** for scalable big data processing pipelines, implementing **MLOps** practices to streamline the deployment and monitoring of machine learning models.
- Created dynamic data visualizations using **Tableau** for data-driven decision-making and developed responsive frontend interfaces with **React** and **JavaScript** to enhance user experience and data visualization.
- Utilized **SQL** for complex backend data management and querying, implemented **Git** version control, and established **CI/CD** pipelines for streamlined code management, automated testing, and continuous deployment.

**Data Scientist** May 2019 – Apr 2021

*Pago Analytics, India*

- Utilized **Python** and deep learning libraries to design resume **parsing** techniques to build a Talent Acquisition Management System (TAMS). Developed advanced **AI** features with 30% more efficiency in the hiring process.
- Engineered an automated hiring process and facilitated its deployment using **AWS Sagemaker**. Collected 10,000+ resumes and labeled them atomically and manually for the training data, using libraries like **NLTK**, **NLP**, and **Spacy** to analyze data.
- Normalizing **SQL** databases, improving data handling and system performance by 2X. Coordinated with cross-functional teams throughout the development and deployment process, following **Agile** methodologies.
- Applied strong analytical skills and creativity in data mining and defining essential **KPIs/metrics**. Communicated complex data points in a comprehensible manner to stakeholders using **Power BI**. Fostered lasting relationships, ensuring smooth project execution, stakeholder satisfaction, and participation in code review.

## PERSONAL PROJECTS

**RAG Loan Document Processor**

- Created **RAG** pipeline with **AWS Bedrock embeddings**, **ChromaDB**, and multi-LLM support (GPT-4/Claude-3) for loan doc analysis, implementing caching and exponential backoff to achieve 99% uptime.
- Optimized PDF/CSV processing via chunking, batch embedding, and rich terminal interfaces, cutting processing time for 10k+ documents, includes reference document name along with lines no for validation.

**Multi LLM Auto Web Scraping Agent**

- Developed LLM-driven web scraper using **LangChain**, Gemini 2.0 Flash-Lite, and **Playwright** to auto-generate/test scraping code.
- Built interactive UI with token tracking, leveraging **BeautifulSoup** for structural analysis and error-refinement loops to reduce manual scripting by 100%.

## SKILLS

| | |
|---|---|
| **Languages:** | Python, R, JavaScript, SQL, NoSQL. |
| **Data Science:** | Machine Learning, Big Data, OpenCV, LLM Agents, RAG, MLOps, A/B testing, Fine-tuning Models. |
| **Packages/Tools:** | Scikit-Learn, Graph neural network, NumPy, SciPy, Pandas, NLTK, LangChain, Matplotlib, Keras, TensorFlow, PyTorch, GIT, JIRA, FAST API, React, Cloud Computing, Hugging Face Transformers, BERT. |
| **Databases:** | MySQL, MongoDB, Cassandra, Elasticsearch, PostgreSQL, DynamoDB, AWS RDS, Vector databases. |
| **Statistics/ML:** | Statistical Analysis, Predictive Analytics, Statistical Modelling, Linear/Logistic Regression, K-means Clustering, Classification, Data Mining, Decision Trees, Random Forests, XG-Boost, LGBM, SVM. |

## EDUCATION

**BOWLING GREEN STATE UNIVERSITY, OHIO** Aug 2021 – May 2023

*M.S in Data Science*

- Capstone Research Project: Graph Neural Network-Based Anomaly Detection in Multivariate Time Series Sensor Data.

## CERTIFICATIONS

- **Coursera:** Python and Machine Learning, Data Analytics.
- **Udemy:** Tableau, PowerBI, Django.
- **Data Camp:** Introduction to R programming and statistics in spreadsheets and Excel.