**ANURAG KALAPALA**

anuragworkprof@gmail.com | +1(978)-943-2495 | [LinkedIn](https://www.google.com/search?q=https://www.linkedin.com/in/anuragkalapala) | Dallas, TX

**SUMMARY**

Highly analytical and results-oriented **Data Scientist** with a Master's degree in Computer Science and over 2 years of demonstrated experience in **applied machine learning modeling**. Strong fundamentals in **Probability, Statistics, and Linear Algebra**, coupled with expertise in **Natural Language Processing (NLP)** and **Deep Learning**. Proficient in **Python** and **SQL**, with practical exposure to **PySpark** and **Hadoop/Hive** for analyzing **data at scale**. Proven ability to design, implement, and integrate **AI models** into production, adhering to **best practices in model development**, **unit testing**, **CI/CD**, and **creating well-maintainable & tested codebases**. A collaborative team player passionate about solving real-world business problems and delivering actionable, data-driven insights.

**EDUCATION**

**MS in Computer Science** | University of Massachusetts Lowell                                                    Jan 2023 - May 2024

* **Quantitative Coursework:** Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Data Mining
* **Focus:** Algorithmic solution implementation, model evaluation, data exploration and analysis.

**BE in Electronics and Communication** | Chaitanya Bharathi Institute of Technology           June 2018 – May 2022

* **Coursework:** Data Structures, Digital Communications, Signal Processing, Statistics

**PROFESSIONAL EXPERIENCE**

**Machine Learning Engineer** | HealthEdge                                                                                 July 2024 - May 2025

* Built and integrated **predictive AI models** for patient event forecasting using **Python**, TensorFlow, and scikit-learn, demonstrating applied machine learning modeling for customer behavior.
* Developed a modular MLOps pipeline using MLflow for **experiment tracking and model versioning**, with automated **CI/CD via GitHub Actions** and AWS CodePipeline, ensuring a **well-maintainable & tested codebase**.
* Deployed ML services as containerized microservices (Docker + FastAPI) for scalable, **highly performant, real-time predictions**, optimizing key decisions.
* Implemented **real-time model monitoring** with Prometheus and Grafana, enabling drift detection and continuous performance **evaluation** to elevate model performance.

**Graduate Research Assistant** | University of Massachusetts Lowell                                     June 2023 - May 2024

* Developed **deep learning models** for enterprise text analytics using TensorFlow and **PyTorch**, applying **Natural Language Processing (NLP)** techniques.
* **Fine-tuned BERT and RoBERTa models** for domain-specific named entity recognition, reducing error rates by 30% and improving **data extraction efficiency and understanding**.
* Conducted **research** into advanced AI applications, contributing to the understanding of complex **algorithms** and their practical implications, and **collaborated** on interdisciplinary projects.

**RPA Developer** | Genpact                                                                                            June 2021 – December 2022

* Developed **Python automation scripts** for **data analysis** and created a COVID-19 trend prediction pipeline using **scikit-learn (applied machine learning modeling)**.
* Performed **social media scraping and sentiment analysis**, demonstrating skills in **data collection, exploration, and analysis** from unstructured sources, and producing **actionable insights**.

**TECHNICAL PROJECTS**

**Enterprise Misinformation Detection Platform** | Python, BERT, NLP, Hugging Face, Docker, FastAPI

* Engineered an **end-to-end NLP pipeline** for classifying 17,500+ articles, demonstrating expertise in **data exploration, analysis, and applied machine learning modeling**.
* **Fine-tuned BERT/SBERT** models and deployed as containerized microservices, showcasing **highly performant code** and adherence to **CI/CD best practices**.
* Scaled architecture for high throughput on AWS ECS, demonstrating ability to handle **data at scale** and apply **foundational programming principles**.

**AI-Powered Visual Classification System** | PyTorch, ResNet-18, Computer Vision, MLflow

* Built a full ML pipeline for dermatoscopic **image classification (Computer Vision)**, including **data wrangling, model training, and validation**.
* Tracked model performance via **MLflow** and deployed as a REST API, emphasizing **model evaluation** and a **well-maintainable codebase**.
* Implemented **CI/CD** for training and deployment using **GitHub Actions** and **Dockerized workflows**, adhering to **foundational programming principles** and **unit testing** concepts.

**TECHNICAL SKILLS**

**Quantitative & Statistical Analysis:** **Probability Theory, Statistics, Linear Algebra**, Statistical Learning, Optimization Theory, Time Series Analysis (Conceptual), **Regression, Clustering, PCA, Forecasting**. **Data Science & Machine Learning:** **Applied Machine Learning Modeling**, Deep Learning, **NLP**, Computer Vision, **GenAI Techniques**, AI Models, Predictive Modeling, Model Design, Fine-tuning, Evaluation, Experimentation, Feature Engineering.

**Programming & Tools:** **Python (Proficient)**, **SQL (Proficient)**, **PySpark (Exposure/Eager to Learn)**, **Hadoop/Hive (Exposure/Eager to Learn)**, Java, TensorFlow, PyTorch, scikit-learn, Pandas, NumPy, MLflow, Hugging Face Transformers, Git, Docker.

**Data & Databases:** **Data Exploration & Analysis**, Data Wrangling, **Data Visualization**, Relational Databases (MS SQL, Postgres, MySQL concepts), NoSQL (MongoDB).

**Software Engineering Practices:** **Unit Testing**, **Code Organization**, **CI/CD**, **Documentation**, Prototyping, **Data Structures**, **Algorithms**, Agile Processes, Scalable Systems, Microservices.

**Communication:** Data-Driven Storytelling, Visualizations, Graphs, Narrative Development.

**CERTIFICATIONS**

**DeepLearning.AI Deep Learning Specialization** | TensorFlow Developer Certificate – Google | **Generative AI with LLMs – DeepLearning.AI** | **LangChain for LLM Application Development – DeepLearning.AI** | Automation Anywhere RPA Professional Certificate | UiPath RPA Developer Foundation