Akhil Uthappa

akhilus@bu.edu | (+1)857-498-3232 | LinkedIn | Github | Website

PROFESSIONAL SUMMARY

Data scientist with a proven track record in optimizing data frameworks and developing high-accuracy machine learning models. Skilled in reducing costs and processing times, enhancing AI functionalities, and managing cross-functional teams. Experienced in statistical analysis, testing methodologies, and data-driven decision-making. Expertise in healthcare, e-commerce, and urban development projects. Strong leadership and project management abilities, driving impactful projects to successful completion. Passionate about working towards a sustainable future for the ecosystem.

WORK EXPERIENCE

Data Science Fellow | **Civera** *Mar 2024- Present* **Boston, MA**

- Optimized Civera's data framework with Spark and AWS, reducing processing times by 40% and costs by 30%. Conducted feature engineering and visualized data, boosting transparency in nine states.
- Developed a YOLO ensemble model with Google Vertex AI Vision, achieving 95% accuracy in redacting PII and increasing sales by 40%. Curated ML workflows with MlFlow, Kubernetes, and Docker.

Generative AI R&D Associate | Cleveland Clinic *Feb 2024 - Present* **Boston, MA**

- Developed and optimized a Med-PaLM model with Retrieval-Augmented Generation (RAG) to answer queries about academic research papers. Optimized model accuracy using Hugging Face Transformers, and a vector database with FAISS for similarity search. Leveraged Elasticsearch for data retrieval.
- Trained the model with parallel training on BU's Shared Computing Cluster, using A100-80G GPUs with CUDA to accelerate LLM training & development.

Data Science & Engineering Intern **| Staples Inc.** *May 2023 - Sep 2023* **Boston, MA**

- Developed and optimized regression algorithms on AWS clusters for predictive analytics, enhancing data-driven decision-making processes
- Collaborated with the data engineering team to refine data preparation for machine learning models, leading to a 20% increase in model efficiency
- Implemented a context-aware text-to-SQL RAG LLM service using the Spider dataset contributing to a 45% reduction in query generation time

Data Science Researcher **| Mass General** *Jan 2023 - Jul 2023* **Boston, MA**

- Reduced OPEX by 30% using predictive models for breast cancer data, improving treatment personalization, & enhanced data processing with Spark & Redshift.
- Applied statistical analysis and ML algorithms to the datasets, resulting in a 40% increase in data accuracy & a 25% enhancement in patient info integrity,

Data Science Research Assistant & Fellowship for Technical Research **| Boston University** *Sep 2022 - Present* **Boston, MA**

- Led a cluster analysis of M&A transactions over 250 million USD, resulting in a publication at EAI CSECS 2023 & revealing key data trends & industry patterns.
- Engineered predictive models that significantly amplified user engagement by 20% for a pioneering VR collaborative painting application, leveraging C#, Unity, and OpenXR, which expedited the product's market launch and established new benchmarks in user experience excellence.

Software Engineer II- Machine Learning **|DeepVue.tech** *Apr 2022 - Jul 2022* **Bengaluru, India**

- Collaborated with the data engineering team to optimize data pipelines, automating ingestion & transformation processes, and reducing manual handling by 30%.
- Managed scalability and performance of microservices, initiating services and CI/CD pipeline on Google Cloud and GitHub Actions.

Technical Project Manager- SWE II **| Verloop.io** *Jan 2020 – Mar 2022* **Bengaluru, India**

- Implemented robust data processing solutions with Python, TensorFlow, and Apache Spark for NLP tasks like sentiment analysis and text classification.
- Optimized PostgreSQL & MongoDB databases, reducing data retrieval times by 35%, & enhanced AI chatbot functionality with SpaCy & TensorFlow, increasing CSAT scores by 40%.
- Managed 3 software engineers along with multiple projects and effectively collaborated in a dynamic, cross-functional environment., reducing the development-to-deployment time by 60%

SWE I, **DesignString** *Dec 2018 - Jan 2020* **Bengaluru, India**

- Developed an aviation asset management platform using off-the-shelf ML services, Django, and PostgreSQL and contributed to Frontend with ReactJS
- Automated database backup & encryption processes using Python & OpenSSL with S3 for secure cloud storage, resulting in a 50% reduction in manual workload

LEADERSHIP EXPERIENCE

Talkvert, ***Co-founder and Chief Technical Officer*** ***Mar 2021 – May 2023 Bengaluru, India***

- Co-founded & led the development of a virtual stage platform with ML capabilities using Kotlin & Flutter, managing cross-functional teams & overseeing UI/UX.

Captain, Indian National Soccer Team

- Captained the Indian national team at an international tournament in Beijing, China, achieving runner-up status after playing for a top club in India.

Volunteer, One Community Global **Jan 2024 – May 2024, San Gabriel, CA**

- Conducted operations, code PR reviews, and project management for sustainable city development projects, contributing to innovative urban solutions.

SKILLS

Methodologies and Technologies: OO Design Patterns, Agile/Scrum Development, Scrum, Agile, JIRA, Git, Unity, Open XR, VR Development, A/B testing
Programming Languages: Java, Python, C++, HTML/CSS, JavaScript (Node.js, React), R
Machine Learning & Analytics: Unsupervised (K-means, PCA, etc.) and Supervised (Nearest Neighbor, Naive Bayes, Decision Trees, Linear Regression, SVM), Neural Networks, GAN, LLMs, Predictive Modeling, TensorFlow, PyTorch, Scikit Learn, Apache Spark, Hadoop, MlFlow, MlOps, Hugging Face
Web & Cloud Technologies: Django, FastAPI, Flask, RESTful API, SOAP API, AWS (EC2, EMR, RDS, S3, Lambda, Redshift, Athena, Glue), Databricks/Spark, PostgreSQL, SparkSQL, Python, Pandas, PySpark, Apache Airflow, Google Cloud(Compute, Dataproc, BigQuery, Vertext AI), Azure, Docker, Kubernetes
Databases & others: MySQL, PostgreSQL, MongoDB, Tableau, Power BI, Excel

PROJECTS

E-commerce ETL Data Warehouse

- Developed a robust e-commerce data warehousing solution using Python for data cleaning/ pre-processing, PostgreSQL for database management, pgAgent for job scheduling, and Tableau for insightful data visualization and reporting

Distributed Deep Learning using TensorFlow

- Implemented a distributed DC GAN model on multiple clusters using TensorFlow for generating unique portraits, utilizing AWS SageMaker for scalable cloud computing and distributed training techniques to optimize performance

Analyzing Remote Work Productivity

- Conducted a comprehensive R-based study on remote work's productivity impact due to covid19, employing classification algorithms like Rpart, J48, KNN, Gradient Boosting, and Naive Bayes, along with a range of attribute selection techniques including RFE, Gain Ratio, PCA, and Information Gain, for a significant 70% accuracy.

BU CollaborART

- Developed a collaborative art VR game using C#, OpenXR, and Unity, identifying artist issues in existing applications with Agile methodology

EDUCATION

Boston University, *MS in Computer Science and Information Systems with Data Science specialization* **2022-2024 Boston, MA**
Coursework: Advanced Machine Learning, Big Data & Analytics, Advanced DB Management, Data Mining, Data Science with Python, Distributed Systems, IT Strategy & Mgmt

Visvesvaraya Technological University, *Bachelor of Engineering, Computer Science* **2014- 2018 Bengaluru, India**
Coursework: Data Structures & Algorithms, Machine Learning, OOP, Software Engineering, Computer Networks, Database & Mgmt

PUBLICATIONS & AWARDS

- Published author for EAI Conference on Computer Science and Education in Computer Science EAI CSECS 2023.
- Published research in eM&Ai, the European M&A Institute.
- Runner-Up representing India in the Great Wall Cup International Soccer Championship held in Beijing, China

INTERESTS

Snowboarding, Sailing, Soccer