

# Akhil Theerthala

Bengaluru | akhiltvs@gmail.com | 8179528501 | LinkedIn | GitHub | HuggingFace

## Summary

Senior Data Scientist with 2 years of FinTech experience, specializing in data-centric methods for resource-efficient LLMs. Proven community impact through open-source personal finance models, with community-adapted versions attracting over 25,000 downloads. Achieved significant business results, including a 97.5% reduction in document processing time and a 27.6% improvement in table detection performance.

## Experience

**Research Volunteer** | *Financial Services Innovation Lab, Georgia Tech*

August 2025 – Present

- Developing generalized financial reasoning models and validating their performance through rigorous ablation studies across multiple model scales and difficulty bands.

**Senior Member Data Scientist** | *Perfios Software Solutions*

April 2025 – Present

- Benchmarked Vision Language Models for table structure recognition, yielding a 0.85 TEDS score on bank statements.
- Engineered a document legibility framework that reduced analysis time for documents with unreadable text by 20%.
- Designed and prototyped multiple agentic workflows for underwriting, establishing a proof-of-concept for new products.

**Member Data Scientist** | *Perfios Software Solutions*

June 2023 – April 2025

- Reduced document processing time by 97.5% (from 8s to 200ms) by implementing and optimizing a multimodal classifier.
- Boosted table detection performance by 27.6% by engineering a new module and evaluating multiple model architectures.

## Projects

**Kuvera: A Data-Centric Personal Finance LLM & Dataset**

HuggingFace Dataset

- Engineered a cost-effective pipeline to generate a 20k-sample Personal Finance Chain-of-Thought (CoT) dataset, attracting over 1,400 downloads.
- Leveraged this custom CoT dataset to fine-tune 8B/14B parameter LLMs for efficient single-GPU deployment, resulting in models with over 500+ community downloads.

**Identifying Social Media Engagement Drivers for CrowdFunding projects**

GitHub

- Engineered an NLP pipeline to extract features from Facebook posts, including sentiment, category, and engagement metrics.
- Achieved a 0.82 F1 score with an ensemble model in predicting engagement for social media posts.

## Education

**Indian Institute of Technology, Kharagpur** | *B. Tech, Aerospace Engineering*

Aug 2019 – May 2023

**Related Coursework:** Graphical & Generative Modelling for ML, Dependable & Secure AI-ML, DL Foundations, AI Foundations

## Publications

**Akhil Theerthala** (2025). "A Data-Centric Framework for Training Behaviour-Aware Personal Finance Language Models." *FinNLP Workshop at EMNLP*.

## Certifications

Generative AI Nanodegree (*Udacity*) | Deep Learning Nanodegree (*Udacity*) | Machine Learning in Production (*Coursera*) | The Reasoning Course (*HuggingFace*)

## Skills

**ML & Deep Learning (Python):** PyTorch, HuggingFace (Transformers, TRL), TensorFlow, Scikit-learn, NLTK, OpenCV

**Data Science & Visualization:** Pandas, NumPy, Matplotlib, Seaborn, Plotly, Scrapy

**Developer Tools:** Docker, Git, Tensorboard, Selenium

## Achievements

- Awarded 1st Place in the Reasoning Dataset Creation Competition, a global challenge hosted by HuggingFace, Together.ai, and Bespoke Labs.
- Selected as a Finalist in the Perfios IdeaFest, a company-wide innovation competition, for pitching an AI-powered healthcare product.