**myfi**

# MyFi Hackathon Challenge

## 🧠 Hackathon Challenge : LLM, Find My Fund!

## 🎯 Problem Statement

Most FinTech platforms today manage massive datasets of Stocks, Mutual funds, ETFs, and other securities—especially in the Indian market. Yet, their search capabilities are limited, often relying on databases of fuzzy string match systems.

These traditional methods struggle with incomplete inputs, typos, and ambiguous queries.

Your mission: Build a compact, intelligent language model that accurately matches a user's query to the right mutual fund or security—even with fuzzy, partial, or unclear inputs.

———

## 💡 What You'll Be Solving

Given the securities data:

1. A dataset of Indian security names (stocks, mutual funds, ETFs)
2. Metadata: fund house, category, sub-category, asset class, holdings, sector etc.

Your system must take user queries like:

- "Icici infra"
- "SBI tech growth"
- "High return funds"
- "Get Tax funds"
- "Funds with hdfc holdings"
- "Get funds from tech sector"
- "Get funds whose AUM is greater than 1000cr"

And resolve them to the most relevant and accurate fund/security name.

———

## 🔍 Challenge Highlights

Go beyond traditional search.

# myfi

## MyFi Hackathon Challenge

You'll need to:

1. ❌ Avoid relying on string matching or SQL fuzzy queries
2. ✅ Build or fine-tune a Small Language Model (SLM) or compact LLM (≤7B params)

## Key Requirements:

🧠 Smart disambiguation using metadata (e.g., fund house, sector)

⚡ Optimize for latency—model should work in real-time systems

🔁 Handle typos, partial names, abbreviations, and ambiguous phrases

———

## 📦 Dataset Provided

Find below data for

1. [Mutual Fund Data](Mutual Fund Data)
2. [Stock Data](Stock Data)
3. [Mutual Fund Holding Data](Mutual Fund Holding Data)

———

## 📊 Goals

Goal Description

1. 🔍 High Accuracy - Resolve fuzzy/ambiguous queries to correct fund search
2. 🧠 Metadata Usage - Use category/type info to boost matching
3. ⚡ Low Footprint - Keep model small (deployable on commodity infra)
4. 🧪 Final Evaluation - Performance on unseen test queries
5. 📜 Explainability - Explain how the model made its decision

———

# myfi

## MyFi Hackathon Challenge

## 🛠️ Tools You Can Use

Models:

TinyLlama, DistilBERT, MiniLM, BGE, Instructor, etc.

Libraries

Sentence Transformers, HuggingFace, PEFT/LoRA, QLoRA, OpenVINO

Techniques

- Embedding-based search (Vector DB + reranker)
- Metadata-enhanced retrieval
- Lightweight fine-tuning
- Quantization + deployment (e.g., BentoML, ONNX)

———

## ✅ Deliverables

1. ✏️ GitHub Repo with:
   a. Source code
   b. Setup instructions
2. 🧠 Trained model (or retraining instructions)
3. 📊 Writeup covering:
   a. Model architecture + size
   b. Metadata usage strategy
   c. Evaluation metrics and results
4. 🖥️ Demo (Video/Live): (CLI or UI optional)
   a. Walkthrough of your system handling real-world queries

Let the hunt for LLM, Find My Fund! begin. Good luck, and happy hacking! 💡🧠📊