CS5800 Summer 2025: Project Proposal
Aria Druker, Linghe Zhou

Context and Personal Relevance

Investing can be overwhelming for beginners due to the vast number of financial products available. With thousands of investment options available, selecting the right assets that align with one's financial goals, risk tolerance, and time horizon is challenging. Exchange-Traded Funds (ETFs) are popular investment vehicles because they offer diversification and lower costs, but selecting the right ones depends on an individual's financial goals, risk tolerance, and investment timeline.

This project seeks to simplify the investment selection process by developing a personalized ETF recommendation system. Users will answer a short questionnaire about their financial preferences, and the system will analyze key metrics from 169 iShares ETFs to suggest the top 5 most suitable funds for their profile.

Personal Relevance:

This project is personally significant to me because I have always been interested in personal finance and algorithmic decision-making. By combining finance with programming, I can create a tool that not only enhances my technical skills but also provides practical value to users looking to make informed investment decisions. (Vanessa)

Financial freedom is a personal goal for me and many of my friends. We know that achieving it requires careful planning and optimizing our savings. However, with so many investment opportunities, we often feel overwhelmed by the options available. I hope this tool will help identify ETFs that best align with both my friends' and my financial goals. (Aria)

Scope of the Project

The project will focus on:

- Developing a questionnaire (3 questions, multiple choice) to assess a user's financial goals, risk tolerance, and investment horizon.
- Using the Yahoo Finance Python library to extract relevant ETF data (e.g., compound annual growth rate, volatility standard deviation, max drawdown, Sharpe ratio)
- Implementing a scoring or ranking system to match user profiles with suitable ETFs.
- Recommending the top 5 iShares ETFs based on the user's inputs.

What the project will NOT cover:

- Non-iShares ETFs (scope limited to iShares for simplicity).
- Continuous spectrum of user profiles for preferences
- In-depth financial analysis (e.g., management fees or market predictions)
- Suggestions for real-time trading and balanced portfolio management.

Plan of Action

Week 1: Research & Questionnaire Design

- Finalize the list of user profile questions (e.g., risk tolerance, investment timeline, annual return goals).
- Research iShares ETFs and identify key metrics for evaluation (e.g., standard deviation, historical annual returns, Sharpe ratio).
- Use yfinance to fetch historical ETF data.
- Clean and structure the data for analysis.

Week 2: Data Collection & Processing

- Develop a 2D or 3D graph-based scoring system (e.g., x = risk, y = return).
- Calculate and place user profiles and ETFs in the graph.
- Build a baseline model recommendation logic using brute force, calculating the minimum distance between the user profile and all ETFs.

Week 3: Algorithm Development & Testing

- Build a KDTree structure from the ETF points on the graph to enable pruning of branches, so only viable candidates for the 5 closest ETFs are evaluated.
- Test the system by using differing windows of time from training to testing and see how 5 recommended ETFs performed vs 5 ETFs within a similar range of risk rating.

Week 4: User Interface & Finalization

- Develop a simple user interface.

Data

The primary data source will be:

- **Yahoo Finance API (yfinance)** – Used to extract historical performance, volatility, and other key metrics for iShares ETFs.
- **iShares ETF List** – A predefined list of 169 ETFs to filter and analyze.