George Soros Biography ChatBot & Intelligent Pairs Trading System

Team: George Soros
Members: Hitesh, Meet, Vinay
Course: CSYE 7380 Theory and Practical Applications of AI Generative Modeling

Project Overview

Two-Part Intelligent Financial System

Part 1: Soros Biography ChatBot

- AI-powered Q&A system about George Soros
- Transformer/RAG model with ChatGPT fallback

Part 2: Pairs Trading Strategy Bot

- Automated pairs trading analysis
- Real-time stock fundamentals
- ChatGPT-powered trading insights

Inspiration - George Soros

Why George Soros?

- Legendary investor and philanthropist
- Pioneer of reflexivity theory in markets
- Founder of Quantum Fund
- Famous for "breaking the Bank of England" (1992)
- Net worth: ~$8.6 billion (donated $32B to charity)
- Philosophy: Markets are inherently unstable and predictable patterns emerge from chaos

Project Objectives

Key Goals

- **Educational Tool**: Learn about Soros' investment philosophy through AI
- **Practical Application**: Implement pairs trading - a market-neutral strategy
- **AI Integration**: Combine traditional finance with modern AI/ML
- **User-Friendly Interface**: Make complex trading strategies accessible
- **Real-Time Analysis**: Live market data integration

Part 1 - Soros ChatBot Architecture

Technical Implementation

- **Knowledge Base**: Curated biography data
- **Model Choice**: Paraphase-MiniLm-L6-v2
- **Preprocessing**: Text chunking, embedding generation
- **Retrieval Method**: FAISS vector search over curated documents and chunked answers
- **Fallback System**: Seamless ChatGPT-3.5 integration for low confidence
- **Response Attribution**: Clear source identification

Data Preparation for ChatBot

Biography Data Processing

- **Source Materials**: Books, articles, interviews about Soros
- **Data Structure**:
  - 4000+ Questions and answers pairs
  - Key life events with dates
  - Investment philosophy quotes
  - Major trades and outcomes
- **Quality Control**: Manual verification of facts
- **Format**: Structured Excel for easy updates

NLP Model Implementation

- Fine-tuned encoder on Soros Q&A pairs (MiniLM/BERT family) for domain-specific semantics.
- Supervised training on question–answer pairs with cosine similarity loss for sentence embeddings.
- Context window optimization: chunk long answers; keep prompts within model token limits to avoid degradation.
- Performance metrics: similarity at k, hit rate on seen/near-seen queries, latency per stage (encode, retrieve, generate).

ChatGPT Integration

- **Fallback Intelligence**
- **API Integration**: OpenAI GPT-3.5/4
- **Prompt Engineering**:

"You are an expert on George Soros. Answer based on his known biography and philosophy. If uncertain, acknowledge limitations."

- **Response Handling**: Async processing
- **Cost Management**: Token optimization

Part 2 - Pairs Trading Overview

Market-Neutral Strategy

- **Concept**: Profit from relative price movements
- **Key Principle**: Mean reversion between correlated assets
- **Risk Management**: Market-neutral positioning
- **Soros Connection**: Reflexivity in market relationships
- **Implementation**: Statistical arbitrage approach

Pairs Trading Mathematics

- **Core Statistical Concepts**
  - **Cointegration Test**: Johansen/Engle-Granger

    Tests long-term equilibrium relationship
  - **Z-Score Calculation**:

    Z = (Spread - Mean) / StdDev
  - **Trading Signals**:
    - Long A/Short B when Z < -2
    - Short A/Long B when Z > 2
    - Close positions when |Z| < 0.5

Stock Selection Interface

- **User-Friendly Design**
- **Predefined Pairs**:
  - AAPL/MSFT (Tech Giants)
  - JPM/BAC (Banking)
  - XOM/CVX (Energy)
  - [7 more pairs]
- **Custom Selection**: Manual ticker input
- **Validation**: Real-time ticker verification
- **Sector Matching**: Automatic sector alignment check

Fundamental Analysis Display

- **Comprehensive Stock Information**
- **Per Stock Display**:
  - Company name and sector
  - P/E Ratio (valuation metric)
  - Market capitalization
  - Trading volume (liquidity)
  - Dividend yield
  - Beta (market correlation)
- **Data Source**: yfinance API
- **Update Frequency**: Real-time on selection

Trading Analysis Pipeline

1. Download Historical Data (1-2 years)
2. Calculate Price Spread
3. Test for Cointegration
4. Compute Rolling Statistics
5. Generate Z-Score Series
6. Identify Entry/Exit Points
7. Backtest Strategy
8. Calculate Performance Metrics

Technical Challenges & Solutions

- **Data Quality**:
  - Challenge: Incomplete biography data
  - Solution: Multi-source validation
- **Model Accuracy**:
  - Challenge: Context understanding
  - Solution: Hybrid approach with fallback
- **Real-time Processing**:
  - Challenge: API latency
  - Solution: Asynchronous operations
- **Cointegration Stability**:
  - Challenge: Relationship breaks
  - Solution: Rolling window analysis

Future Enhancements

Roadmap for Improvement

- **Enhanced NLP**:
  - Multi-language support
  - Voice interaction
- **Advanced Trading**:
  - Machine learning for pair selection
  - Dynamic position sizing
  - Options strategies integration
- **Additional Features**:
  - Portfolio management
  - Risk analytics dashboard
  - Mobile application

Conclusion & Q&A

- **Project Summary**
  - Successfully integrated AI with quantitative finance
  - Created educational tool honoring Soros' legacy
  - Demonstrated practical application of pairs trading
  - Built user-friendly interface for complex strategies
- **Thank You!**

  Questions?