Comprehensive Guide to AI for Alpha Generation

1 Phase 1: Foundations - Essential Finance and AI Concepts

1.1 Module 1.1: Core Financial Concepts and Intuition

This module introduces fundamental financial concepts crucial for AI-driven alpha generation, including Alpha, Beta, Efficient Market Hypothesis (EMH), and the Capital Asset Pricing Model (CAPM). It emphasizes conceptual understanding and practical calculations of Beta and expected returns.

1.2 Module 1.2: Setting Up Your Local Python Environment

Focuses on the creation and configuration of an optimized Python environment suited for finance and AI projects, covering essential libraries like NumPy, Pandas, Matplotlib, scikit-learn, yfinance, and statsmodels. Includes instructions for creating structured directories and managing virtual environments.

1.3 Module 1.3: Accessing and Understanding Market Data (Numerai API)

Explores accessing financial datasets through APIs, specifically Numerai. This module addresses downloading datasets, understanding data structure, exploring and preprocessing datasets, and understanding the use-cases for Numerai's anonymized financial market data.

1.4 Module 1.4: Simple Backtesting with Zipline or Backtrader

Introduces backtesting frameworks Zipline and Backtrader, covering installation, basic concepts of backtesting, strategy implementation, and results interpretation. Special emphasis on analyzing strategy effectiveness and adjustments for improved performance.

2 Phase 2: Supervised Learning and Factor Models

2.1 Module 2.1: Feature Engineering for Finance

Covers the critical aspects of creating informative financial features from raw data. Techniques such as creating lagged returns, volatility measures, momentum indicators, and economic indicators are explored.

2.2 Module 2.2: Linear Regression and Factor Models

Introduces factor models like the Fama-French three-factor model, covering theoretical foundations, factor selection, and the implementation of regression models to explain and predict returns.

2.3 Module 2.3: Regularization Techniques (Lasso, Ridge)

Explores regularization techniques to prevent model overfitting. Discusses theory behind Lasso and Ridge regression, their implications, and practical implementation guidelines for financial data.

2.4 Module 2.4: Factor-Based Stock Selection Strategy

Guides students through developing and implementing a systematic factor-based stock selection strategy. Covers factor selection, backtesting procedures, and portfolio performance evaluation.

3 Phase 3: Deep Learning for Time-Series Forecasting

3.1 Module 3.1: Understanding Time Series Data

Introduces fundamental concepts such as stationarity, trends, seasonality, and techniques to preprocess and analyze financial time series data effectively.

3.2 Module 3.2: LSTM Networks for Stock Price Prediction

Detailed exploration of Long Short-Term Memory (LSTM) neural networks for capturing temporal dynamics in stock price data. Includes theory, architecture design, hyperparameter tuning, and model evaluation.

3.3 Module 3.3: CNNs for Pattern Recognition in Financial Data

Introduces Convolutional Neural Networks (CNNs) and their applications in detecting and leveraging complex patterns in financial time series. Includes data preparation, architecture design, training strategies, and practical considerations.

3.4 Module 3.4: Practical Deep Learning: Model Training and Optimization

Discusses CPU-optimized methods for deep learning model training, parameter optimization, and techniques to ensure efficient model performance and reliability on local computational resources.

4 Phase 4: Reinforcement Learning and Algorithmic Trading

4.1 Module 4.1: Introduction to Reinforcement Learning

Provides a solid theoretical grounding in reinforcement learning, including Markov Decision Processes (MDP), Bellman equations, value functions, and policy optimization.

4.2 Module 4.2: Implementing Simple Trading Policies (Q-learning)

Demonstrates practical implementation of Q-learning for creating simple algorithmic trading policies, including strategy design, state/action definition, reward structures, and policy evaluation.

4.3 Module 4.3: Policy Gradient Methods and Advanced RL Strategies

Introduces advanced reinforcement learning algorithms like policy gradients, Actor-Critic methods, and Deep Reinforcement Learning for trading. Emphasizes theoretical foundations, implementation strategies, and potential pitfalls.

4.4 Module 4.4: End-to-End Trading Bot Prototype

Guides the development of a fully autonomous trading bot using combined RL strategies. Includes the integration of data pipelines, real-time execution, backtesting, and evaluation metrics to finalize a deployable prototype.