Quantitative Algorithmic Trading System

Student Name: Liu Chen
Student ID: 22100974D
Programme: BSc (HONS) COMPUTER SCIENCE

Supervisor: Prof. Henry C. B. Chan
8 October 2025

Contents

1 Background and Problem Statement **2**

2 Objectives and Outcome **2**

2.1 Objectives 2

2.2 Expected Outcome 3

3 Project Methodology **3**

3.1 Phase 1: Data Processing and Model Development 3

3.2 Phase 2: Strategy Backtesting and Evaluation 4

3.3 Phase 3: System Integration and Prototyping 4

3.4 Literature Review 4

4 Project Schedule **4**

4.1 Semester 1:Modeling and Backtesting (Sep 2025 - Jan 2026) 4

4.2 Semester 2: Automated Trading System Design and Evaluation (Feb 2026 - May 2026) 5

5 Resources Estimation **5**

6 References **6**

1 Background and Problem Statement

The financial markets, particularly the equity markets, are complex systems driven by a multitude of factors. The Efficient Market Hypothesis (EMH) suggests that asset prices fully reflect all available information, making it impossible to consistently achieve returns in excess of average market returns[1]. However, empirical evidence has shown that short-term inefficiencies and predictable patterns, often termed "anomalies," do exist, especially in high-frequency data [2].

The rise of computational power and data science has led to the proliferation of algorithmic trading, where computer programs are used to execute trades at speeds and frequencies impossible for a human trader. This project focuses on Hong Kong-listed securities, with primary analysis on liquid instruments such as the **Tracker Fund of Hong Kong (2800.HK), stocks, or derivative warrants**. The 2800.HK ETF, which tracks the Hang Seng Index, serves as the primary research subject due to its high liquidity and extensive tick data availability.

Problem Statement: The core problem this project aims to solve is to investigate whether modern machine learning techniques, specifically time-series models, can be applied to the high-frequency tick data of the 2800.HK ETF to identify and capitalize on short-term market inefficiencies. The challenge lies in developing a model that can generate predictive signals with enough accuracy to be profitable after accounting for real-world trading frictions. While 2800.HK serves as the primary case study, the methodology is designed to be generalizable to other liquid Hong Kong-listed securities including individual stocks and warrants

2 Objectives and Outcome

The primary goal of this project is to design, implement, and evaluate a complete data-driven algorithmic trading system. The scope focuses primarily on the 2800.HK ETF, with the framework designed to be extensible to other Hong Kong-listed securities such as individual stocks or derivative warrants

2.1 Objectives

The project is broken down into three specific and measurable objectives:

1. **Develop Predictive Models:** To apply and evaluate various machine learning and time-series models on high-frequency data on Hong Kong-listed securities (primarily 2800.HK, with potential extension to individual stocks or warrants) to forecast short-term price movements or trends. For example Tree model: XGBoost[3], the basics RNN: LSTM[4], Transformer-based model: Informer[5], MASTER[6], Graph Neural Networks: HIST[7] or even try to gain feature information from large language model FinGPT[8].

2. **Construct a Backtesting Framework:** To build a robust historical simulation (backtesting) framework in Python. This framework will be used to systematically assess the performance of trading strategies derived from the predictive models.

3. **Implement a Prototype System:** To engineer a prototype of an automated trading system that integrates the entire workflow, from data ingestion and processing, through model prediction, to the generation of simulated trade orders.

2.2 Expected Outcome

Upon successful completion, this project will deliver:

- A set of trained predictive models for the 2800.HK ETF.
- A comprehensive backtesting report detailing the profitability and risk profile (e.g., Sharpe Ratio, Maximum Drawdown) of the developed strategies.
- A well-documented Python codebase for the entire trading system prototype, including modules for data handling, modeling, and simulation.
- A final project report summarizing the methodology, findings, and potential avenues for future research.

3 Project Methodology

The project will be executed in three distinct phases, corresponding to the objectives outlined above. This phased approach allows for iterative development and evaluation.

3.1 Phase 1: Data Processing and Model Development

This phase focuses on data and modeling.

- **Data Acquisition and Preparation:** High-frequency tick data for 2800.HK and potentially other liquid Hong Kong-listed securities (e.g. derivative warrants) will be acquired. The raw data will be cleaned, processed, and resampled into appropriate time bars for feature engineering.
- **Feature Engineering:** A variety of features will be engineered from the price and volume data, such as moving averages, volatility measures, RSI, MACD, VWAP and Bollinger Bands.
- **Model Training and Selection:** Several models will be trained on the historical data to predict daily price direction in the next week into with three classes: up significantly,down significantly, or not determined. The models' predictive power will be evaluated using standard classification metrics on a held-out test set.

3.2 Phase 2: Strategy Backtesting and Evaluation

This phase validates the model's signals in a simulated trading environment.

- **Framework Implementation:** An event-driven backtesting engine will be built in Python to avoid lookahead bias.
- **Strategy Logic:** The predictive signals from Phase 1 will be translated into concrete trading rules (e.g., "buy if prediction is 'up significantly').
- **Performance Analysis:** The strategy will be backtested over a long historical period. Key performance and risk metrics will be calculated to judge its viability.

3.3 Phase 3: System Integration and Prototyping

This final phase combines all components into a cohesive system.

- **System Architecture:** A modular system will be designed, separating concerns for data, strategy, execution, and portfolio management.
- **Implementation:** The components will be implemented in Python, creating a simulation engine that can "run" on historical data as if it were live.

3.4 Literature Review

The project's methodology is grounded in a review of relevant academic and practical literature. Foundational concepts of systematic strategy development are drawn from works like Chan [2]. The predictive modeling component will survey the evolution of time-series forecasting techniques, starting with established models like XGBoost [3] and LSTMs [9], which serve as strong baselines. I will then investigate state-of-the-art, deep learning architectures including specialized Transformers like Informer [5] and MASTER [6], and explore Graph Neural Networks (HIST) for capturing inter-market relationships [7]. Additionally, the potential of using financial Large Language Models (FinGPT) for novel feature extraction will be considered [8]. This review directly informs the choice of models and the critical design of a robust backtesting framework to ensure valid performance evaluation.

4 Project Schedule

The project schedule is structured over two semesters (Sep 2025 - May 2026)

4.1 Semester 1:Modeling and Backtesting (Sep 2025 - Jan 2026)

- **Data, Features, and Baseline Models (Sep - Oct 2025):** Acquired and processed high-frequency 2800.HK data. Engineered features and established performance benchmarks with baseline models (e.g., XGBoost, LSTM).

  - **Milestone 1:** A clean, feature-rich dataset is prepared and baseline model performance is documented.

- **Advanced Modeling and Backtesting (Nov 2025 - Jan 2026):** Develop a robust, event-driven backtesting engine. Implement, train, and evaluate advanced models (e.g., Transformer, GNN) to identify the optimal predictive model for the system.

  - **Milestone 2:** A functional backtesting framework is complete. A superior predictive model is selected based on rigorous comparative analysis.

4.2 Semester 2: Automated Trading System Design and Evaluation (Feb 2026 - May 2026)

This semester is dedicated to the research, design, and implementation of a complete, end-to-end automated trading system prototype.

- **System Architecture and Implementation (Feb - Mar 2026):** Research best practices for trading system design. Implement a modular architecture that integrates data ingestion, model-based signal generation, risk management, and simulated order execution logic.

  - **Milestone 3:** A prototype of the end-to-end automated trading system is implemented, capable of running a full historical simulation from raw data to performance report.

- **Comprehensive Evaluation and Finalization (Apr - May 2026):** Conduct extensive simulations with the prototype system. Perform sensitivity analysis on transaction costs and slippage. Write the final report, document the system architecture and codebase, and prepare the presentation.

  - **Milestone 4:** The final project report, fully documented codebase, and presentation slides are submitted, detailing the system's design, functionality, and performance.

5 Resources Estimation

The project will rely on standard, readily available resources.

- **Hardware:** A standard personal computer with a multi-core CPU and at least 8GB of RAM. Access to a consumer-grade NVIDIA GPU based on Google Colab.

- **Software:** All software is open-source. The project will use Python 3.x with standard data science libraries (Pandas, NumPy, Scikit-learn, Matplotlib) and a deep learning framework PyTorch.

- **Data:** Historical high-frequency (tick) data for 2800.HK and selected Hong Kong-listed securities will be purchased

- **Personnel:** The project will be undertaken by me, with regular guidance and feedback from the academic supervisor Prof. Henry C. B. Chan.

6 References

[1] Eugene F Fama. "Efficient capital markets: A review of theory and empirical work". In: *The journal of Finance* 25.2 (1970), pp. 383–417.

[2] Ernest P. Chan. *Algorithmic Trading: Winning Strategies and Their Rationale*. Hoboken, New Jersey: John Wiley & Sons, 2013.

[3] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[4] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[5] Haoyi Zhou et al. "Informer: Beyond efficient transformer for long sequence time-series forecasting". In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. 12. 2021, pp. 11106–11115.

[6] Tong Li et al. "Master: Market-guided stock transformer for stock price forecasting". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 38. 1. 2024, pp. 162–170.

[7] Wentao Xu et al. "Hist: A graph-based framework for stock trend forecasting via mining concept-oriented shared information". In: *arXiv preprint arXiv:2110.13716* (2021).

[8] Xiao-Yang Liu et al. "Fingpt: Democratizing internet-scale data for financial large language models". In: *arXiv preprint arXiv:2307.10485* (2023).

[9] Thomas Fischer and Christopher Krauss. "Deep learning with long short-term memory networks for financial market predictions". In: *European journal of operational research* 270.2 (2018), pp. 654–669.