Hybrid Model Time Series Prediction Strategies

A Quantitative Analysis Platform for Combining Statistical and Machine Learning Models

Adam Kuzniewski
akuzni22@student.aau.dk

Benjamin Olsson Høj
bhoj22@student.aau.dk

Emil Suphi Doganci
edogan22@student.aau.dk

Mathias Mosgaard Larsen
mmla22@student.aau.dk

Maximilian Marin Duhn Gotthardsen Scarpa
mscarp22@student.aau.dk

Rasmus Damsboe Kragelund
rkrage22@student.aau.dk

***Abstract*—This paper concerns the modular construction of strategies for time series prediction. These modules include basic logic components along with statistical models and machine learning models. The purpose of centered strategies around a hybrid approach is to evaluate the usefulness of this form of time series prediction.**

I. Introduction

Time series prediction has become a critical component of various modern analytical and decision making systems. Whether the task is to forecasts weather or energy demand, anticipate component failure in mechanical systems, or understand and predict financial market dynamics, the ability to model future data is a major advantage in most domains, and improves system performance and informed decisions. Regardless of the application, the efficiency of forecasting and decision making depends on methods capable of capturing structure in temporal data.

Recently, machine learning models have begun to compete with classical statistical and quantitative models as methodology for time series prediction. Traditional quantitative strategies rely on technical analysis to identify historical patterns through statistical indicators [**?**]. Although interpretable, these linear methods often fail to capture complex market dynamics. In contrast, deep learning models, particularly Long Short-Term Memory (LSTM) networks, excel at modeling non-linear dependencies [**?**]. However, standalone machine learning models frequently suffer from overfitting and lack of transparency in decision-making [**?**].

“Hybrid strategies” that integrate deterministic technical indicators with probabilistic machine learning predictions offer a potential solution to these limitations [**?**]. Such strategies aim to take advantage of the strengths of both domains and address the shortcomings of each approach in isolation, allowing construction of models that are interpretable and accurate across various time series application domains.

To facilitate the investigation of the efficacy of using hybrid strategies for time series modeling, a modular platform was developed that integrates LSTM inference directly into a strategy builder. This system allows for the composition of logical dataflows in which machine learning predictions serve as auxiliary signals alongside traditional quantitative models within a broader technical framework.

II. Related Work

The engineering of automated trading systems has historically presented a high barrier to entry, requiring a synthesis of advanced software architecture, low-latency data processing, and complex domain logic. Traditionally, translating a trading hypothesis into a functional system involved writing extensive boilerplate code to manage data feeds, synchronization, and order execution [**?**]. This technical debt often creates a disconnect between the design of a strategy and its implementation. Consequently, the field has seen a shift toward visual programming and low-code solutions, aiming to abstract these engineering complexities into modular, reusable components [**?**].

In terms of strategy logic, the field is generally divided into two forms. The first relies on technical analysis, where traders use mathematical formulas like Moving Averages or RSI to find trends. These methods are transparent and easy to understand but can be rigid [**?**]. The second form utilizes machine learning, particularly deep learning models like LSTMs, which are excellent at finding hidden patterns in data but often act as ”black boxes” that are hard to interpret [**?**].

Recent research suggests that the most effective approach is not to choose one over the other, but to combine them. By using simple logic rules to filter or confirm the complex predictions made by AI models, traders can create ”hybrid” strategies that are both powerful and safer to deploy [**?**]. However, most existing software platforms force users to pick a side: they are either simple rule-based builders or complex code-first ML frameworks. There is a lack of tools that seamlessly integrate these two paradigms in a visual, user-friendly manner.

III. Preliminary Work

For the purposes of this project, the focus was primarily aimed at analysis of financial market data. While multiple domains were considered for evaluating hybrid time series prediction strategies, financial market data was ultimately selected due to practical reasons. A large amount of high-quality historical price data is mostly freely available from various sources, and this accessibility and reduced data acquisition challenges made it a primary choice.

The aim of the project, however, is not necessarily limited to financial analysis. Various types of time series can be passed to constructed hybrid models, regardless of their domain of origin. While some models may specialize in certain domains, or be trained to predict more accurately on certain markets, the fundamental principle of hybrid strategies will likely carry over to other domains.

The decision to utilize Long Short-Term Memory (LSTM) networks over more contemporary architectures, such as Transformer-based models, was primarily driven by the balance between predictive performance and implementation complexity. Although newer models may have offered improvements in accuracy or more interesting results, they would have most likely required significantly larger datasets, computational resources, and more specialized knowledge to implement and train properly. For the scope of this project, the primary research objective is the evaluation of the hybrid strategy framework itself, rather than the development of novel deep learning architectures. LSTMs showcase a well-established standard in time series forecasting with extensive documentation and library support, which would allow quick integration into the platform and ensure that development efforts remain focused on the system's modularity and the interaction between the machine learning and quantitative components.

IV. Problem Statement

Composition of effective and accurate time series forecasting and prediction strategies is difficult, partially due to the large number of approaches and tools available, including classical statistical methods, indicator-based quantitative models, and the more modern machine learning approach. Each tool provides certain advantages, but also comes with each their disadvantages. To maximize the gain of the advantages of each of the selected methodologies, while minimizing or alleviating the disadvantages, a way to thoughtfully and fruitfully structure hybrid strategies would possibly provide this quality.

Hybrid strategies have emerged as a promising methodology for time series prediction and forecasting, but as of yet remain source? largely unutilized . Without structure, however, strategies become difficult to compare, gauge, and generalize. In order to compose and uniformly structure these strategies, it is advantageous to consider them directed acyclic graphs (DAG), where nodes correspond to transformations or decisions based on the information flow of the edges between nodes of the graph. Furthermore, constructing strategies as DAGs also allows for sequential execution of strategies, enabling frequent and precise evaluation of strategy accuracy on certain sets of data.

Concisely, the problem statement for this project should reflect the ability to structure strategies in a modular manner, in which the composed strategies may utilize hybrid methodologies, and evaluate their performance during sequential execution over a set of data. Thus, the project aims to design a modular framework for composing structured hybrid time series prediction strategies, consisting of deterministic and machine learning components, and evaluating their performance through sequential execution on historical data.

V. Method

Working with financial market data requires reliable data providers with high-quality data available. While many provider options are available, the primary considerations were Yahoo Finance, EODHD and Massive (formerly Polygon.io) for stock data, along with Coinbase and Binance for crypto currency data. For crypto currency data, the most consistent dataset, at least for Bitcoin (BTC), was sourced from Coinbase. Regarding stock data, the selected provider was Massive, as they provide access to consistent historical data across the American stock market. Due to both providers being free to use, along with their consistent data quality, made them the selected providers for the crypto currency market and stock market, respectively.

Since financial time series data is often represented through the Open-High-Low-Close (OHLC) format, and the aim of the project is to make a general purpose time series prediction framework, this format must be generalized. To represent financial market data as general purpose time series, each OHLC entry is split in multiple series, where each serie only has a single vector of data points.

VI. Contribution

The project encompasses a central API that handles data-ingestion API kommunikere med DB's igennem localhost. alt andet kommunikere med API'en. API'en bruger OpenAPI standarden, som bliver rendered med Scalar

VII. Conclusion