Enhancing Investment Recommendations with Temporal Learning Models: A Study on Stock and ETF Portfolios

MSc AI for Business
Practicum 2

Caio Cesar Teixeira de Lima
Student ID: 23354135

School of Computing
National College of Ireland

Supervisor: Dr Muslim Jameel Syed

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Caio Cesar Teixeira de Lima
Student ID:	23354135
Programme:	Practicum 2
Year:	2025
Module:	MSc AI for Business
Supervisor:	Dr Muslim Jameel Syed
Submission Due Date:	05/08/2025
Project Title:	Enhancing Investment Recommendations with Temporal Learning Models: A Study on Stock and ETF Portfolios
Word Count:	10389
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th September 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Investment Recommendations with Temporal Learning Models: A Study on Stock and ETF Portfolios

Caio César Teixeira de Lima
23354135
National College of Ireland

Abstract

The use of technologies and application of artificial intelligence in investment management has advanced significantly in recent years, but we still used legacy systems, often assume static investor profiles, ignoring temporal changes in behavior and market conditions. This research proposes the use of temporal models such as LSTMs and HMMs, to develop a system of dynamic investment recommendations based on historical data on stocks and ETFs, seeking to fill a gap in the continuous adaptation of portfolios and the validation of their financial impact. The motivation arises from the need for tools that respond to stock market volatility and offer operational accuracy to investors in an automated and controllable way. The methodology includes time series analysis, model training, and validation with financial metrics in simulated portfolios. It is expected to contribute to an accessible and scalable framework, providing adaptive recommendations and performance evidence.

1 Introduction

The introduction of AI technologies applied to personal finance and investment management is enabling a new era in how financial decisions are made, with greater accuracy and efficiency compared to traditional approaches. The stock market stands out as an area of growing interest, given its dynamism and relevance to individual and institutional investors. However, a persistent behavior in automated portfolio management is the ineffectiveness of many systems in adapting investment recommendations to continuous changes in investors' financial behavior and market conditions. Since traditional models assume static risk profiles, disregarding the temporal evolution of these variables, the inherent volatility in the stock market requires more flexible approaches. This research seeks to address this challenge by exploring the potential of temporal models such as Long Short-Term Memory (LSTM) neural networks and Hidden Markov Models (HMMs), to create dynamic recommendations based on historical data stock and ETFs.

As one of the most demanded points in this study area is the creation of financial solutions that combine automation, customization, and real-time adaptation. These characteristics are essential given the context of globalized market, subject to rapid transformations in which we inserted. Investors, whether amateur or professional, face difficulties in manually adjusting their strategies to market fluctuations, while portfolio managers seek tools that optimize returns and minimize risks efficiently. In evidenced by Pangavhane et al. (2023), the author suggests that modern technologies can transform investment management, but significant gaps remain. The literature review points out that, although studies such as Ambuli et al. (2024) demonstrate the potential of machine learning techniques to improve predictive accuracy in portfolios, there is a lack of systems that integrate dynamic adaptation and practical validation in real market scenarios. Shahani et al. (2023) highlights the use of sequential models, such as LSTMs, a theme that we will address in this study, to monitor changes in financial behavior, but their practical application in the stock market is still limited, as well as the proof of their financial benefits.

This study proposes to solve the problem of staticity in investment recommendations by developing a framework based on temporal models that use data within the scope of stock histories and ETFs. The research problem motivates the following question:

> "How can time-learning models, using historical stock data and ETFs, improve the adaptability and accuracy of investment recommendations and empirically validate their impact on portfolio performance in the stock market?"

The proposed solution is based on training models such as LSTMs and HMMs to capture temporal patterns in price and volume series, generating recommendations that evolve with market conditions. These recommendations will be tested in simulated portfolios, validated by financial metrics such as the Sharpe Ratio and the Maximum Drawdown, offering practical evidence. The expected result is the delivery of an affordable and scalable system that is capable of technical prediction and practical impact, benefiting investors and financial platforms such as robo-advisors.

The author structured this document in six sections, where the first introductory session presents the research problem, its relevance and motivation for its issue, establishing a practical and theoretical context for this study, following the second session, where in the literature review explores research in artificial intelligence and finance, highlighting the gaps that we seek to address on dynamic adaptation in investments and financial impact validation. The third, session of methodology, describes the research in general, its approach, including strategic dates, model criteria, experimentation, and design. The fourth session outlines the technical frameworks, tools and architecture used and proposed in the solution, including data resources, pipelines and how to configure the model, and visualizations and setups. The fifth evaluation section presents experiments conducted to compare models and their performances, analysing the results in different scenarios. And the conclusion, session 6, summarizes everything that was found, outlines limitations and suggestions for future work.

2 Related Work

The use of techniques and tools with artificial intelligence in the financial market has grown rapidly and is positioned due to the great ease of data volume treatment, identification of linear parameters in high volatility environments. Neural network models, such as LSTM, computational networks and, more recently, transformer-based architectures, have added value to the objective of predicting prices, detecting demands and detecting anomalies in the market as a quantitative negotiation strategy. The authors were able to analyse several studies in which models based on Deep Learning presented results superior to traditional approaches such as ARIMA and linear regression for time series prediction, study Ambuli et al. (2024).

Therefore, it is observed that, in most of the research, the focus is on predictive accuracy. This ends up not connecting these results to a practical and real strategy for portfolios allocation or real investment decisions. Additionally, few studies consider market regime changes such as high volatility environments or moments of stability and instability. In the decision-making process, some studies explore the use of hidden Markov models, HMM, to identify these regimes. But in general, this approach is not integrated into prediction models based on Deep Learning, which limits its practical application.

The authors of the research sought to fill this gap together, proposing a hybrid architecture in combination with a time series model, LSTM, and the probabilistic identification of regimes, the HMM, in order to generate signals for purchase and sale and for portfolio allocation. This proposal differs from the others in delivering market forecasts with direct allocation actions, and accounts for the dynamic nature of a market, adapting to decisions according to changes in the financial context.

2.1 Public data sources with Yahoo Finance

The authors of this research chose to use Yahoo Finance as a public data source due to its history and reputation. They were able to analyse several recent studies in which financial data is used in predictive analysis models. For research requiring predictive analysis, Yahoo Finance has been adopted due to its data capillarity, historical series, and consistency, especially when integrated with learning models such as the LSTM proposed in this research. With free data availability via API and Python library, its widespread use is reinforced in the study Komosar and Kijanovic (2023). The authors propose a critical discussion of the definition of a financial data source as well as its availability in different formats. In this scenario, Yahoo Finance stands out for its institutional reliability and purpose. The study Yan et al. (2021) highlights the search for transparency of economic data in automated modeling pipelines, proposing a multidimensional assessment framework that includes data quality and sources, highlighting factors such as reliability, stability, and integrity. Although the study does not primarily focus on Yahoo Finance, its criteria were aligned with the data availability provided by the tool, corroborating its suitability for data-intensive use. According to the study Dhuhita, Wibisono, Widodo and Pambudi (2023), it explicitly uses historical gold price data from Yahoo Finance between 2000 and 2023 to train an LSTM model to simulate the one we will use in our research, achieving a remarkably low RMSE of 0.000033. Daily resolution and completeness of the data set allowed for a robust hyperparameter tuning, and this experience proves the value of the platform for long-term commodity trend modeling. The study Ferdiansyah et al. (2019) the same time, Yahoo Finance is also used to predict the price of the Bitcoin for five years. Their LSTM model, evaluated with an RMSE of 288.59, outperformed traditional approaches such as SVM, reinforcing that, even in highly volatile markets, Yahoo Finance provides stable and consistent historical data. Preis et al. (2013), which, although it does not concentrate on time series modeling, investigates a user behavior dataset on Yahoo Finance itself. Their research does so by showing that browse activity on the site can forecast the volume of stock transactions up to three days, indicating Yahoo Finance provides not only the numerical data but also behavioral signals with predictive power outperforming traditional models for web searches on both accuracy and timeliness. In the end, Daita et al. (2022) built a high performance cryptocurrency forecasting model using LSTM, obtaining training data directly from Yahoo Finance via API. Their method focuses on the daily closing prices and moving averages of major cryptocurrencies, pursuing 98% accuracy and an RMSE of 0.26. The good performance of the model, emphasizes the stability and granularity of Yahoo Finance data, which is necessary to train deep learning networks with low noise. In these studies, we find that Yahoo Finance is an established and flexible data source that fulfills critical research requirements with respect to accessibility, formatting, update intervals, and empirical soundness, providing a definitive benchmark for this research.

2.2 Processing of financial data

For normalization purposes, particularly Minimax and Scale, appears as a core technique in several studies dealing with financial magnitudes. This can be observed in studies Wong et al. (2023), Seshagani et al. (2025), and Reddy et al. (2024),. The authors emphasize minimax normalization as a single path for scale input, facilitating conversion and improving machine learning techniques in distinct models such as XGBoost, CNN-RNN, or hybrid models. These studies highlight that the normalization of financial series, such as stock prices, in technical indicators, brings better interceptions of machine learning layers, especially when the inputs come from multiple resources, several units and orders of magnitude.

For the preparation of data timing, which is another key layer of the pre-processing process, financial models depend not only on values but on the order and pace of values over time. Works Wong et al. (2023), Reddy et al. (2024), and Pradhan et al. (2023), explicitly deal with the organization of data into time windows, as well as sliding sequences and horizons. This strategy ensures that time dependencies are preserved and is a critical mechanism for prediction models such as LSTM or CNN-LSTM. In some cases, such as citation Wong et al. (2023), metrics like the average test time are used to measure your responsiveness, bringing a temporal arrangement in a delivery time that affects forecasts in possible real-time.

The revised literature brings together and consolidates solid foundations for our choice of Data Preparation. When we are in a concept of Machine Learning or in a classic algorithm, validation is important in the identification of outliers, normalization, the Inputs, Distributions and the Base Time structure are fundamental steps for the construction of solid models, mainly when we deal with financial data, which requires adaptability and accuracy depending on market conditions.

2.3 Time Series Prediction with LSTM Networks

The studies that deepen the use of Long Short-Term Memory (LSTM) networks have highlighted their ability to model time dependencies in financial series. As we can see in the study Kumar et al. (2023), the authors apply LSTM to the SENSEX index, integrating analysis of news sentiment, surpassing the ARIMA in short-term forecasts. The inclusion of qualitative data adds agility, but the sensitivity to the number of times suggests being unstable. In the study Ferdiansyah et al. (2019), focused on Bitcoin (2014-2019), the LSTM reaches an RMSE of 288.59, surpassing SVM, but fails to capture extreme volatility due to external events. Compared to Kumar et al. (2023), which explores synergy with feelings, Ferdiansyah et al. (2019) relies more on historical data, showing the need for greater resilience.

The study Dhuhita, Al Farid, Yaqin, Haryoko and Huda (2023) the authors showed results related to the price of gold with RMSE of 0.000033, using performance improvement for hyperparameters, but its performance is limited by unmodelled macroeconomic factors. Similarly, Daita et al. (2022) achieves 98% accuracy in cryptocurrencies but is vulnerable to sudden changes, while Ambuli et al. (2024) predicts stocks like Google with 99.46% accuracy, surpassing Linear Regression, although the result could lead to high operating costs. These positive results contrast with the one raised in the study Ambuli et al. (2024), which compares LSTM and ARIMA at SSE 50 (MAPE of 0.76% vs. 1.13%), highlighting the superiority of the LSTM in non-linear patterns, but reinforcing challenges for large-scale problems already evidenced in other studies. In our research, we seek to integrate continuous temporal learning and validation in multiple contexts to overcome these weaknesses.

2.4 Portfolio Optimization with Deep Learning Models

We observed that the use of portfolio optimization has made advances with the use of deep learning models, which go beyond the fixed rules or historical averages traditionally known; these models have the flexibility to deal with the complexity and volatility of financial markets. In the study Ambuli et al. (2024), the authors explore Deep Q-Networks (DQN) as a deep reinforcement learning technique to adjust portfolios in real time. When tested in dynamic scenarios, the model achieves an impressive annual return of 12.5%, with a maximum drawdown reduced to -5.2%, which highlights its ability to minimize risks while maximizing gains. However, again we are faced with the reliance on robust historical data and the need for computational infrastructure raises concerns about scalability.

On the other hand, we can observe that in the study Day and Lin (2019), the author proposes a robo-advisor that combines multiple algorithms, being LSTM, GRU, XGBoost, and Random Forest with the classic optimization models Black-Litterman and Markowitz, focusing on the ETF market, one of the focuses in our research. Implemented in the context of Taiwan's 0050.TW index, the system achieves an annual return of 12%, surpassing traditional benchmarks and strategies based only on historical data. As a result, the strength of this model lies in algorithmic integration, leveraging the ability of LSTM and GRU to capture time trends and XGBoost to handle complex patterns, while Black-Litterman incorporates market insights. However, the authors recognize that market volatility and the need for constant calibration of models can compromise their long-term consistency, thus being a challenge that in our research we address through detailed validation and adaptive adjustments.

In the study Shahani et al. (2023), with the Investopal system, the authors offer a distinct perspective by proposing a customized financial consulting solution for the Sri Lanka market, using XGBoost for investment forecasts, LSTM for trend analysis, and BERT to process natural language in financial education. This focus on personalization, eliminating the need for human consultants, is an innovative factor in the industry, but its applicability is limited by the regional focus and the lack of empirical validation of suggested portfolios. Compared to Day and Lin (2019), which tests a diversified portfolio in ETFs, Shahani et al. (2023) lacks generalization, although it brings valuable insights on the integration of qualitative data. On the other hand, the study Varadarajan and Priya (2024) expands in scale by exploring machine learning applications in fraud detection, portfolio management and regulatory compliance at a global level. Using algorithms such as decision trees, neural networks, and SVM, the model has the main responsibility of analyzing large volumes of data in real time, thus promoting efficiency in algorithmic trading and risk assessment. Despite its advances, regulatory barriers and reliance on consistent historical data limit its universal adoption, a point that our research considers when proposing more flexible solutions.

Complementing, the study Kour (2024) presents practical cases using machine learning in the financial sector, highlighting implementations in companies such as JPMorgan Chase, PayPal and Mastercard. Supervised and unsupervised learning techniques, such as neural networks and pattern analysis, improve accuracy in credit scoring and fraud detection, while portfolio management gains efficiency with data-driven forecasts. However, the authors point out difficulties of interoperability in the models, which may hinder confidence for practical use by investors and regulators. This recurrent obstacle was also observed in the study Varadarajan and Priya (2024). While Ambuli et al. (2024) emphasizes dynamic adjustments and Day and Lin (2019) algorithmic synergy, Shahani et al. (2023) focuses on personalization, Varadarajan and Priya (2024) on the global scale, and the study Kour (2024) on practical applications, they all face challenges such as volatility, generalization or transparency.

One of the objectives of our research will be to integrate the strengths of these approaches by incorporating adaptive temporal learning to assist robust and adaptable portfolio tools. Our proposal goes beyond including detailed empirical validation, even with limitations such as data dependence and computational costs. The goal is to offer a scalable solution that balances accuracy, efficiency, and applicability in diverse financial contexts.

2.5 Reinforcement Learning and Dynamic Strategies

Studies on reinforcement learning have been established as a powerful approach to creating adaptive financial strategies, allowing dynamic adjustments in response to market conditions. In the study Shukla et al. (2024), the authors apply RL with Markov Decision Processes (MDP) to model sequential decisions in portfolio optimization and algorithmic trading. The approach stands out for its ability to simulate choices in uncertain environments, but remains theoretical, with the authors warning of risks of overfitting and high computational complexity, without robust empirical validation in real markets. Already in the study Ambuli et al. (2024), the use of Deep Q-Networks, also called DQN offers a more practical application, achieving annual returns of 12.5% and reducing the maximum drawdown to -5.2%. Tested in real-time adjusted portfolios, the model demonstrates effectiveness but relies on high-quality historical data and computationally intensive resources, which can limit its scalability.

With a complementary view we analyze the study Johnson-Skinner et al. (2021), which proposes a trading pairs strategy that combines Hidden Markov Models (HMM) with the robust Kalman filter to predict volatility and adjust trading signals in cointegrated stocks. The results of this research show superiority over traditional methods, with the DDIVF-HMM model capturing market time variations and increasing profitability. However, the choice of the so-called ideal number of hidden states remains a technical challenge, compromising its consistency in different scenarios. Finally, the study Ilhan et al. (2023) presents the Markovian RNN, integrating recurrent neural networks (RNN) with HMM to deal with non-stationary environments. Tested on synthetic and real data, the model outperforms ARIMA in metrics such as mean squared error, thus offering flexibility for structural changes in data, but again we found challenges due to the high computational cost, thus being a barrier for practical application.

While Shukla et al. (2024) establishes a conceptual basis for RL, Ambuli et al. (2024) and Johnson-Skinner et al. (2021) prioritize applicability with measurable results, and Ilhan et al. (2023) combines theory and adaptability to a promising combination, but is limited by data dependencies and costs. When we apply our research, we explored this synergy by integrating RL with continuous temporal learning, seeking dynamic strategies that are scalable and empirically validated, overcoming these barriers to offer practical solutions in varied financial contexts.

2.6 External Data Integration and Different Models

On the other hand, the use of sentiment analysis has proved to be an important tool in financial forecasts, capturing qualitative variables that complement traditional quantitative data. In the study Botta et al. (2024), the authors used recurrent neural networks (RNN), LSTM and Transformers (BERT) to analyze Stock Twits postings, achieving 75% accuracy in predicting price trends. This approach explores the impact of feelings expressed in social media on market volatility, but the lack of clarity in the models makes it difficult to adopt them in practical decisions.

The study Shahani et al. (2023) that we have already cited earlier, the authors take a different approach when using BERT to process educational texts and recommend investments in Sri Lanka, combining it with XGBoost and LSTM. Although it personalizes suggestions, its local focus and the absence of experimental validation limit its generalization, contrasting with the wider scale found in the study Botta et al. (2024). On the other hand, the study Daita et al. (2022) applies sentiment analysis to cryptocurrencies like Bitcoin and Ethereum, reaching 98% accuracy with LSTM, while Gill et al. (2023) focuses on Ethereum, achieving 95%. Both highlight the influence of external events such as headlines or regulations on volatility, but sensitivity to these external factors reveals a reliance on robust contextual data that is not always available.

These studies demonstrate the transformative potential of unstructured data, but lack consistent integration with quantitative variables and practical validation to an obvious gap. In our research, we seek to overcome this barrier by combining market data with adaptive temporal learning and detailed empirical validation, seeking more robust and applicable forecasts that capture both market trends and external influences.

2.7 Hybrid Models and Computational Advances

In the observed studies, this approach shows that hybrid models have emerged as a promising strategy that seeks to combine different methodologies, seeking to integrate the efficiency of traditional methods with the predictive capacity of advanced machine learning techniques. We note that this approach is particularly valuable in finance, where the complexity of time series requires solutions that balance accuracy, robustness, and practicality.

In the study Zaychenko and Kuzmenko (2022), the authors propose a GMDH-LSTM hybrid network for short-term forecasts in the NASDAQ Emerging Markets Bond Total Return Index (EMBRI). The model combines the Group Method of Data Handling (GMDH), which automatically selects the most efficient network structure, with LSTM, resulting in lower mean squared error and higher computational efficiency at 1-5 day horizons, surpassing pure LSTM. This focus on reducing computational cost is a practical advantage, but the authors of the study do not explore long-term horizons where traditional LSTM may have superiority, suggesting a limitation in broader financial scenarios.

We also note that the study Wei et al. (2024) seeks to move in this direction by integrating ARIMA and LSTM to predict cargo volumes at 57 logistics centers over 121 days. In this case, ARIMA captures linear and seasonal patterns, while its counterpart LSTM models nonlinear relationships and long-term dependencies, resulting in more accurate predictions for the next 30 days than isolated methods. The result of the model is in the balance between statistical simplicity and computational complexity, but the authors recognize challenges in the diversity of data and the need for parameter adjustments, which can make it difficult to adapt directly to more volatile financial assets.

We can already find a contrast, the study Xie et al. (2019) presents the DCLSTM (Dilated Convolution LSTM), a hybrid model that combines dilated CNNs with LSTM to predict gold ETF prices. When the authors use a tested scenario with historical data, DCLSTM achieves an MSE of 1.92 and a coefficient of determination (R^2) of 0.9171, surpassing traditional methods such as ARIMA, KNN, and Back Propagation Neural Networks (BPNN). The inclusion of dilated CNNs broadens the receptive field of the model, allowing the capture of broader temporal patterns, while LSTM keeps the focus on long-term dependencies. Despite its high precision, the study highlights the need for adjustments in the so-called hyperparameters and the dependence on high-quality historical data, which may limit its application in markets subject to abrupt changes.

In summary, we can compare these approaches because Zaychenko and Kuzmenko (2022) prioritizes efficiency and simplicity, making it suitable for rapid predictions, while the study Wei et al. (2024) offers a useful balance for mixed-pattern series, although less focused on direct finance. On the other hand, we note that the study Xie et al. (2019) stands out for its accuracy in specific assets such as gold ETF, but requires greater adjustment effort. These studies illustrate the potential of hybrid models to overcome limitations of isolated approaches, but also reveal common barriers such as the need for robust data and continuous adjustments. In our research, we seek contributions to develop a system that combines adaptive temporal learning with hybrid models, seeking to balance efficiency and precision. Our proposal aims to integrate external variables and detailed experimental validation, overcoming the limitations of scalability and adaptability observed, to offer a practical solution applicable in various financial contexts.

3 Methodology

The objective of this section will be to detail the proposed methodological approach to answer the research question, outlining the steps necessary to develop and validate a framework of dynamic investment recommendations based on temporal models.

3.1 Research Method

This research proposes a quantitative and experimental methodological approach to answer the question:

> ”How can time-learning models, using historical stock data and ETFs, improve the adaptability and accuracy of investment recommendations and empirically validate their impact on portfolio performance in the stock market?”

The proposed framework will seek to integrate time models specifically LSTMs (Long Short-Term Memory) and HMMs (Hidden Markov Models) in a structured pipeline that processes historical financial data to generate dynamic investment recommendations, adapted to changes in the market. The solution aims to overcome the limitation of static systems, offering a proactive approach that evolves with temporal patterns. The methodology is divided into connected modules in which each one is responsible for a step of the process, from data collection to financial validation:

1. **Data Collection and Integration:** To obtain time series we will extract from Yahoo Finance, covering two different periods: experiment 1 five years (January 2020 to December 2024), experiment 2 seven years (January 2012 to December 2019) capturing daily closing prices, maximum, minimum, and trading volumes of shares (ex: S&P 500 companies like AAPL and MSFT) and ETFs (ex.: SPY, QQQ). This data will be extracted through a simple script in Python, using the yfinance library, which allows direct access to the public API of Yahoo Finance. This period was chosen because it included varied scenarios, such as the pandemic crisis and periods of recovery, offering a robust basis for temporal analysis. In addition, we using yfinance with public data for model replication in any scenario.

2. **Pre-processing and Structuring:** The cleaning and normalization of data ensure the quality of inputs into models. The cleaning initializes the identification and treatment of missing values (e.g., holiday gaps) by linear interpolation, which fills in the missing data based on time patterns, a method suitable for continuous financial series. Outliers will be detected with the IQR method (Interquartile Range), but kept if they represent real market events (ex.: sudden drops), because they are informative for the research context. The normalization will be performed with the Min-Max Scaling technique, adjusting the values to a scale between 0 and 1, preserving the relative relationships between prices and volumes. These methods were selected for their simplicity and effectiveness in preparing financial data for machine learning models, minimizing distortions and ensuring comparability.

3. **Temporal Modeling:** We chose to follow with the development of the models, which will be centered on LSTMs and HMMs to capture market dynamics. LSTMs will be configured with three layers (128, 64 and 32 neurons respectively), using the function of tanh activation in memory cells and sigmoid in gates, with 60-day time windows to predict price trends. The training will adjust hyperparameters such as learning rate (initial 0.001, decaying adaptively) via Adam optimization, using cross-temporal validation to avoid overfitting. In the additional models, HMMs will be implemented with 3 to 5 hidden states (e.g., high, low, and medium volatility), estimated by the Baum-Welch algorithm, to classify market regimes based on returns and volatility. This detailed approach will allow replication as it specifies the architecture and parameters, balancing complexity and predictive capability.

4. **Recommendation Generation:** As a result of this research experiment, we will deliver a recommendation system that will adjust portfolio allocations based on model outputs. LSTM price forecasts will be converted into buy/sell signals (e.g., buy if the forecast price rises 2% in 5 days) using a simple threshold rule, while HMM market states will report risk adjustments (ex.: reduce exposure in high volatility states). These signals will feed an optimization algorithm that recalculates the weights of the hypothetical portfolio (e.g.: 40% SPY, 30% QQQ, 30% AAPL/MSFT), maximizing the expected risk-adjusted return. We chose to implement in Python, given the vast amount of libraries and easy development. Allowing daily or weekly dynamic adjustments, with decision tracking

for further analysis, this level of detail ensures that the process is clear and replicable in different market applications.

5. **Financial Validation:** The results of this solution are evaluated through financial validation, which will test the recommendations in a simulated portfolio. Thus, comparing them with a traditional buy-and-hold baseline, a strategy recognized in the financial literature, obtained by simulating the same initial allocation without dynamic adjustments. To make a simple comparison the method will use backtesting through 6-month and 1-year windows, calculating the Sharpe Ratio, Maximum Drawdown, and cumulative return. The experimental setup assumes a portfolio simulation during the following post-pandemic (2021-2022) period, a low period with high volatility (this would test the robustness), and the period pre-pandemic (2012-2019). The results will be shown in tables and graphs for easy risk-adjusted performance comparison, to offer empirical evidence for the feasibility of the framework.

4 Design Specification

The scope that we define for this research will be the analysis of American stocks and ETFs, we will select resources with great accessibility of historical information. The programming language of this project will be Python given the support of artificial intelligence applications and wide market utilization. To follow flexibility and easy experimentation, the author dedicated it by modules and the architecture of the different items of this application. The separate framework structure consists of different components, such as Data Collection, Process, Temporal Module, Signal Generation, and Portfolio Maker Decision. Each individual module allows us to update, test and expand independently. Therefore, there is a large number of segregated scripts.

For the implementation of temporal models, such as LSTMs, the TensorFlow and Keras libraries will be used, which offer an infrastructure to build and train recurrent neural networks. In addition, the library "hmmlearn" will be used for the construction of HMMs, enabling probabilistic analysis of market states. In addition, the Pandas and NumPy libraries will be used for time series manipulation, which includes prices and trading volumes. The code execution and exploratory analysis will be developed in PY scripts, an environment for code execution but also for the immediate visualization of results and detailed documentation of each step. For visual analysis, the Matplotlib and Seaborn libraries will be used to create market trend charts and performance metrics.

Yahoo Finance will also give us data for research purposes. We chose it for data because of the historical volume and availability of data. This is a public source that delivers a detailed time series of daily closing prices, maximum and minimum, and volumes for trading that are the essential data for experimentation in our study. The chosen timeframes reflect two scenarios. The first scenario is proposed in the initial research from 2020 to 2025, which seeks to highlight a specific market regime, such as the COVID-19 pandemic, and the recovery of this period. It is a phase of general normalization. Additionally, for a second experiment, we expanded the scope of the research, bringing a period of calm in the market, which was from 2012 to 2019, to prove the robustness of the model and prove it with a significantly larger data sequence. This temporal diversity proves the robustness and testing of the system, its adaptability, and possible performance in different times and market condition sensitivities. About the stocks, we chose the traded stocks in the S&P 500 index, one of which would be Apple (AAPL) and Microsoft

(MSFT), and also widely traded ETFs such as SPY and QQQ. These data will serve as the basis for building a hypothetical portfolio, consisting of 5 to 10 assets, whose initial allocation (e.g., 40% in SPY, 30% in QQQ, and 30% in individual stocks such as AAPL and MSFT) will reflect the typical choices of an average investor in the stock market.

In our research, these resources and software tools, combined with a set of financial data, represent an optimal scenario for the development and testing of the framework in line with expected academic and practical objectives. Not only does this infrastructure underpin the temporal analysis proposed, but it also guarantees that results are straightforward to interpret and replicable by peers or that they can be implemented in practical investment settings.

5 Implementation

The development of our research sought to develop a system that proposes dynamic investor recommendations based on temporal learning. For this development, the author sought to organise the code in a modular way, where there are three blocks that represent different functions of the machine learning pipeline. First, the data preparation, followed by the modelling, and finally the decision. Each step was made to first ensure modularization and robustness, facilitating replicability and even future maintenance if necessary, considering real market conditions.

All implementations of these projects were made in the Python language, which, for the author, is familiarity due to the great acceptance of the professional market and academic activity. Especially in projects where there is a need to create statistical modelling or even financial analysis. Finally, different source codes files were generated that can be executed in GeoTerminat Blue or can be run separately, in each step at a time, in an environment where Python programs run.

5.1 Data Collection and Preparation

As mentioned above, the database used for this study was obtained from public data extracted from the Yahoo Finance API through the library `yfinance`. Historical series of prices, closing, opening, minimum, and traded volumes were collected. As the experimental assets we chose SPY, QQQ, AAPL, and MSFT, covering the period from 2020 to December 2024.

For research, during the collection, the data was organized separately, in files by asset and saved in a structured report, to then be carried out the loading with these data. The collection was automated as follows:

```
data = yf.download(tickers=['SPY', 'QQQ', 'AAPL', 'MSFT'],
                   start='2020-01-01', end='2024-12-31')
```

Listing 1: Data Collection by yfinance


```
data = yf.download(tickers=['SPY', 'QQQ', 'AAPL', 'MSFT'],
                   start='2020-01-01', end='2024-12-31')
```


Listing 2: Data collection via yfinance

Already in the pre-processing state, the data underwent a linear interpolation and were normalized based on the Min-Max Scaling technique, as described above.

```
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(df[PRICE_COLUMNS + ['Volume'
    ]])
```

Listing 3: Data Standardization

Following outlier detection, we use an interquartile range method (IQR), but maintain them when associated with real market events. The objective of this research is to preserve the historical completeness of financial data.

5.2 Modelagem com LSTM e HMM

As the core module of this research, we use time modeling with two complementary approaches. LSTM networks for trend forecasting, as described in the project planning, and HMM for classification of volatility regimes of market conditions.

LSTM

The extttLSTMPredictor model was structured with three LSTM stacked layers and intermediate extttDropout layers to mitigate overfitting:

```
model = Sequential([
    LSTM(128, return_sequences=True, input_shape=(60, n_features)
        ),
    Dropout(0.2),
    LSTM(64, return_sequences=True),
    Dropout(0.2),
    LSTM(32),
    Dense(1)
])
```

Listing 4: LSTM network architecture

The model was configured in 60-day input windows, where there is a 5-day forecast horizon. And the model was trained with the ADAM shader. This includes temporal cross-validation and early stopping to avoid overfitting.

HMM

The complementary model extttHMMPredictor, based on extitHidden Markov Models, was used in this case it was used to identify the regimes of market volatility (low, medium, and high) based on the closing price and volume provided by the financial market. The modularization of high volatility asset exposure followed the following logic:

```
if market_state == 'high_volatility':
    adjusted_weights[ticker] *= 0.5
```

Listing 5: Adjustment of risk via volatility regime

The model was trained with the Baum-Welch algorithm, using standardized data via`StandardScaler`.

5.3 Rebalancing and Decision Making

We used the extttPortfolioManager component, which is responsible for applying the signals of the models to the simulated research portfolio. Each week, the LSTM forecasts were converted into a buy signal as described in the code below:

```
if predicted_return >= 0.02:
    signals[ticker] = 'buy'
else:
    signals[ticker] = 'hold'
```

Listing 6: Generation of predictive signal

The signals are weighted with the volatility states detected by the HMM model and used to recalculate the portfolio weight, ensuring that the total sum is 100% of the whole portfolio.

5.4 Financial Validation and Visualization

The final step of our research consists in the validation and visualization of the data. The performance of the system was extitbacktesting and, comparing the dynamic portfolio with the extitbuy-and-hold approach, cumulative return metrics were used extitSharpe Ratio and extitMaximum Drawdown.

As for the visualization, it was done with matplotlib and seaborn, and the results were stored automatically. The following is an example of generating portfolio value charts over time, to give a sample of the search result.

```
plt.plot(daily_values['Date'], daily_values['Portfolio_Value'])
plt.title('Portfolio Value Over Time')
```

Listing 7: Viewing the value of the portfolio

Thus, we were able to clarify through an empirical validation, providing evidence of the potential of the system and its ability to adapt in different market regimes, analyze historical data and optimize the risk-return ratio of a portfolio.

6 Evaluation

The purpose of our research was to create a model in which it would traverse time series through the LSTM with shares or ETFs of the American market, but with a concept of prediction of market analysis described in the HMM. In this session, we seek to demonstrate a comprehensive analysis of the results of the study, as well as the results that it can generate in a positive way, or even weaknesses that the study may present. As described earlier in other sessions, we used a data set during a specific period to be able to bring results of this research. In order to support and have an analysis of the research results with statistical and visual data, we chose to break the evaluation session into different forms and possible visualizations, being: the first a pure analysis of the proposed model, where we will explore the results of LSTM plus HMM. A second session, where we will make a comparison where we do not use the HMM that looks for distinct market movements and we will use only contemporary series such as the LSTM. And finally, a third one, where we take a baseline market model like HBoost and compare with the two. And finally, we created a model that I named as Baseline to prove the effectiveness of temporal series. The third experiment is complementary to experiments 1 and 2. The difference we seek is to give a broader scope of data that was proposed in the first studies and proposed as research material for a period of low volatility and low market risk, where we did not have major financial crises, called the bullish period of the American financial market, which was between 2012 and 2019. We went through a period where there were no crises, such as the 2008 real estate crisis and the 2019 COVID crisis. The idea is that we can show broadly the application in different scenarios and make a comparison model in different scenarios, obviously analyzing the effective results of this portfolio, which is the value of the portfolio, profitability and risk-adjusted return.

6.1 Experiment 1: We use the LSTM+HMM models

The first experiment of the study is based on data previously reported throughout the study. A time series will be giver from January 2020 to December 2024, a period where we had several extreme market regime events such as the COVID-19 pandemic, global economic recovery and great volatility due to these events. Data is taken from Yahoo Finance. We will take a portfolio of shares from Apple, Microsoft, SP 500 ETFs, justified due to high liquidity, representativeness of different sectors and practical relevance for individual and institutional investors. Specifically, we will take the opening and closing price, the day's maximum, and the traded volume. The result of this first experiment, we hope to see data showing the performance of the proposed model.

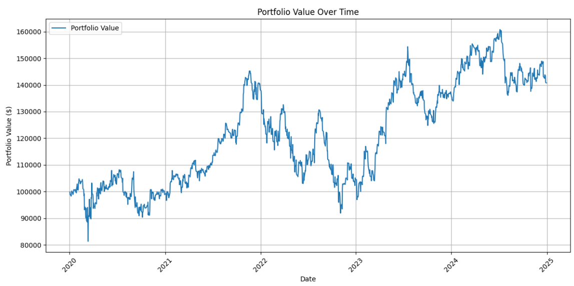

Figure 1: Portfolio perfomance

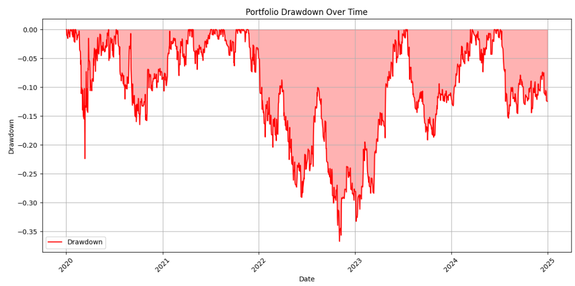

Figure 2: Drawdown over the period

As we can see in the simulated model, the charts have different behaviors, but significantly positive performances. The growth of the portfolio in market value had a strong growth trend. Obviously, in the period of the COVID crisis, the value of the portfolio had a significant fall, the same happens in the Drawdown chart. We can see that the value of assets had a significant drop over the period of COVID, but its performance is positive within what is expected. This shows that the model can be applied to situations of high volatility and market uncertainties.

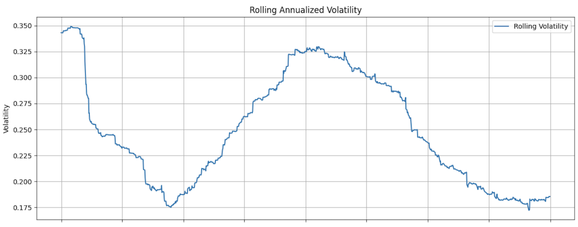

Figure 3: Annualized Volatility

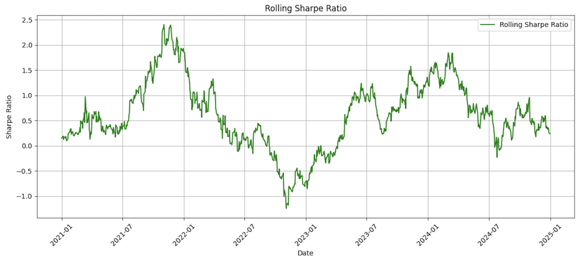

Figure 4: Sharpe Ratio

Again, when we further analyze the graphical distribution of annual volatility and the Sharpe Ratio, we can highlight two important financial indicators, which specifically analyze the risk-return ratio of a portfolio. The annual one measures the variation of the returns of the assets of the portfolio as a whole and follows the same behavior of the value of the portfolio, being that it had the biggest crisis in the period of COVID, and then a recovery, this lowering the volatility, given that there was a stability of the market as a whole. The Sharpe Ratio is the quality of return based on the risk taken, a metric that evaluates the return on investment adjusted for the risk it offers. And then there was a major crisis also in the period of COVID, but it had a degree of stability that goes there between 0 and 0.5, and is permeating this region. This demonstrates, once again, that the model can be met for these market situations, delivering performance, controlled low volatility, and controlled risk.

6.2 Experiment 2: We use the LSTM without HMM model

Our objective in the second experiment is to use the same data that was input in the first experiment, but this time taking the HMM from the model, considering only time series with the LSTM, trying to identify if the performance would be the same when we follow the other indicators presented in the Experiment 1. Thus we can create a comparative degree of performance within the same family of models, where we have time series with LSTM, in the first model following what was proposed in the research, adding a factor that looks at market variations and the moment of crisis, and now without looking at the moment of crisis. And we will make a comparison of the two models to assess their effectiveness and efficiency effective.

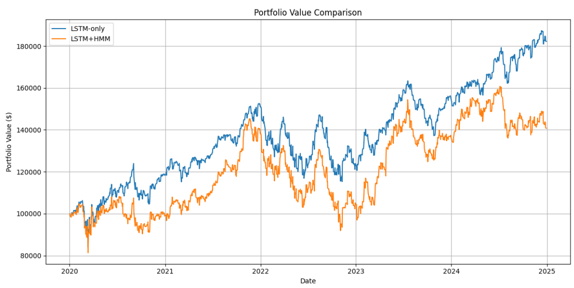

Figure 5: Portfolio Value Comparison

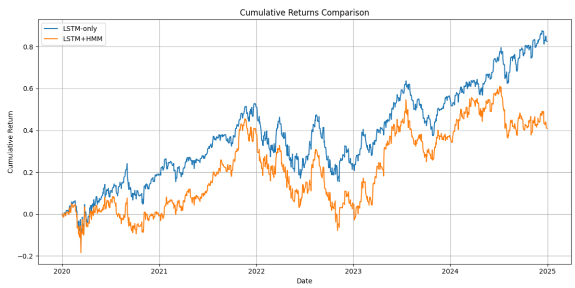

Figure 6: Returns Comparison

When we analyze the results without the HMM model, we can see a significant difference in performance improvement. Both in the value of the portfolio, if you make the comparison of one with another, and in the value of the return, if you make a comparison.

They follow the same market movements, the same sharp fall at the time of COVID, let's say that this market recovery curve follows the same. But the model without HMM, it is effectively more profitable from a profitability point of view.

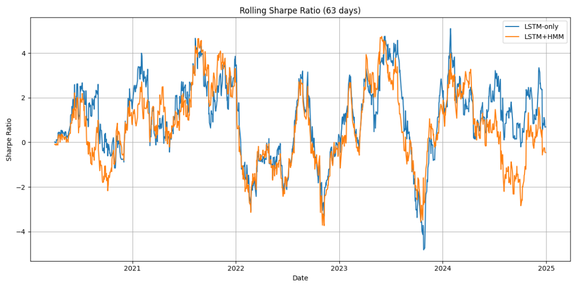

Figure 7: Sharpio Comparison

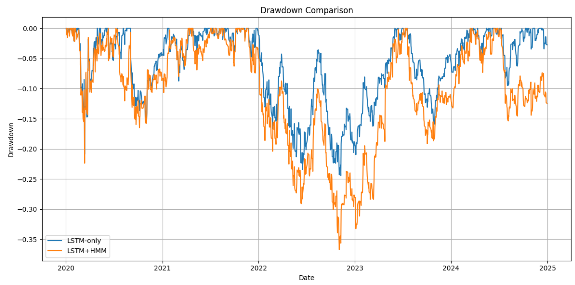

Figure 8: Drawdown Comparison

The result of volatility through Sharpe Ratio and Drawdown, it is very similar to the performance of the portfolio, the value of the portfolio presented in the previous table. The model where HMM is not applied tends to have better indicators, also of volatility and risk-return adjusted.

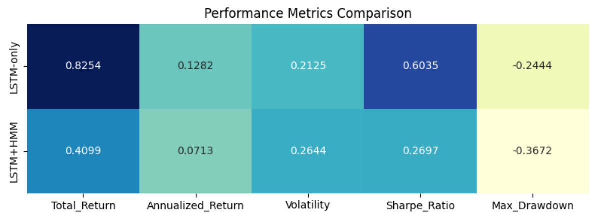

Figure 9: Metrics Comparison

Comparing the performance of both models in a grouped way, we notice that the model without HMM has a total return significantly higher than 0.82 against 0.40 of the model proposed in the research. The analyzed return is also significantly higher than 0.12 versus 0.07. The volatilities are very close, changing from 0.21 to 0.26 in the proposed model. And the Sharpe Ratio is 0.60 versus 0.26. These performance indicators and metrics indicate that the inclusion of HMM for this specific model proposed in the research has an unsatisfactory performance. We will try to clarify this point in the discussion session.

6.3 Experiment 3: We use the LSTM+HMM models data 2012-2019

As mentioned in the introduction, the third experiment is based on experiment 1, where we use the LSTM and HMM models. The difference is that we chose to use a range of dates where there were no financial crises, which changed the dynamics of the market as a whole, from 2012-2019 before COVID. And there were periods of high market to understand what would be the behavior of the model and the effective result in the portfolios, as well as the success metrics of execution of the model, as we did in previous experiments.

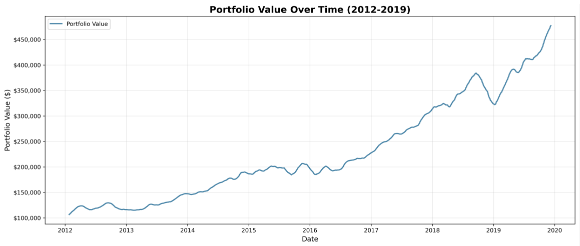

Figure 10: Portfolio Value Over Time (2012–2019)

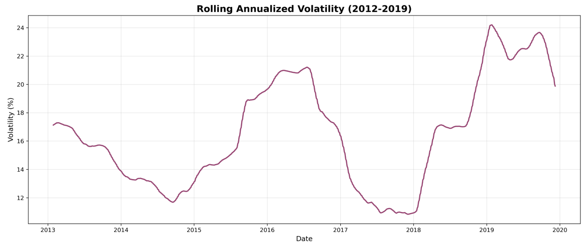

Figure 11: Volatility Over Time (2012–2019)

An experiment applied with another data sampling, larger, and with a period of marked bullish, what we observe are behaviors very similar to the first, of significant portfolio growth. In this case, there was a growth of portfolio value without major downgrades during the period. Volatility was quite high, as there was a period of growth between the years of 2016 and 2017, a fall and a return of growth, up to the end of the pandemic. The model managed, even in a moment of volatility, at a time when there may have been some abnormal event, to maintain a portfolio with exponential growth, following a pattern that had already been modeled in the first experiment with smaller data.

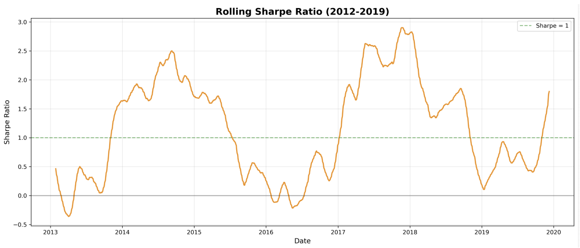

Figure 12: Sharpe Ratio Over Time (2012–2019)

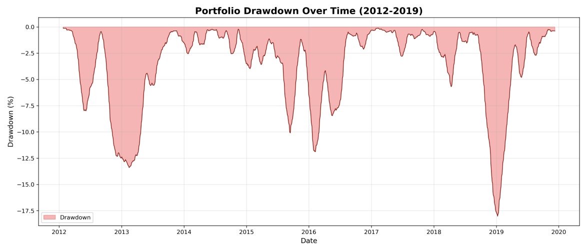

Figure 13: Drawdown Over Time (2012–2019)

The 2012 and 2019 Sharpe Ratio and Drawdown charts help explain the higher volatility observed in Figure 12. Long into the period of the beginning of the 2019 crisis, assets had a large devaluation and a significantly increased return risk. This is due to the fact that the study took data a little beyond the beginning of the 2019 crisis. If you analyze figure 11, the portfolio grew steadily, had a fall due to COVID, but it recovered and managed to follow. This shows the resistance and robustness of the model, being able to make the purchase and sale allocations, and focusing on the continuous growth of the portfolio as a whole.

6.4 Case studies: Models performance

In this case study, we will analyze the performance metrics of the model. Here the author seeks to highlight the metrics and states of the model, where we have the complementary model of LSTM + HMM against only the LSTM. The idea is that we can evaluate the results themselves and be able to take important insights about the result of the experiments and research as a whole. In the table below, the execution results of LSTM+HMM models versus LSTM are shown. The observed states were:

Table 1: Model Evaluation Metrics

Metric	LSTM-Only	LSTM+HMM	Difference
Mean Absolute Error	55.0735	55.4032	+0.3297
Root Mean Square Error	63.9742	64.2832	+0.3090
Mean Absolute Percentage Error	21.9741	22.0996	+0.1255
R^2 Score	0.0920	0.0821	−0.0099
Classification Accuracy	0.5134	0.5171	+0.0037
Precision	0.5052	0.5084	+0.0032
Recall	0.5134	0.5171	+0.0037
F1 Score	0.5055	0.5085	+0.0030

What we can observe, analyzing the results table of the models, is that the LSTM model plus the HMM, it has a small outperformance in the classification metrics. We can observe a gain in recall, accuracy, recall, and F1 score. It performs better, being positive to the result of our research. However, on the other hand, when we look at prediction metrics, they are very low, indicating that both models have a very limited predictive power. And the main squared error is significantly high for both models, for prediction and recall.

- **Model Accuracy Performance:**
  - LSTM+HMM slightly outperforms in most classification metrics
  - Classification Accuracy: LSTM+HMM (51.71%) > LSTM-Only (51.34%)
  - Precision: LSTM+HMM (50.84%) > LSTM-Only (50.52%)
  - Recall: LSTM+HMM (51.71%) > LSTM-Only (51.34%)
  - F1 Score: LSTM+HMM (50.85%) > LSTM-Only (50.55%)
- **Prediction Quality:**
  - R^2 scores are very low (0.0920 vs 0.0821), indicating limited predictive power
  - Mean Absolute Error is similar between models (55.07 vs 55.40)
  - Both models struggle with prediction accuracy

6.5 Discussion

In the discussion session, the author sought to structure in two stages. At first, we look at the result of the proposed study, the proposed model, from the point of view of what was expected, as the management of an investment portfolio. In the second, we segregated by the results of the models, so that we can have a more segmented analysis.

Analyzing purely the information from experiments 1, 2 and 3, where in 1 and 2, which were the initial experiments of the research, and in addition 3, which came to give more robustness to our analysis, it is possible to affirm that the model delivers what was expected. It decreases volatility and enables portfolio reallocation with the constant growth of asset value. Then, both in periods of great market volatility and periods of greater tranquility, the model proved to be efficient through the data that were processed in them. We can analyze from Figure 2 of the portfolio performance that the proposed portfolio, the initial study, it followed a significant growth in a period of crisis. It had a fall that was followed by the entire market, but it had a rapid recovery within this scenario, thus bringing a vision that demonstrates. Yes, the model may be able to deliver forecasts, projections and signals of buying and selling stocks so that an investor, whether institutional or ordinary investor, can have a better risk management of his portfolio as a whole. Showing that what was proposed in the initial research delivers academic value to the line of research on finance and investments.

Additionally, we were able to observe in the experiment where we removed HMM from the proposed model and ran the same data only with LSTM, that HMM reduces portfolio volatility, reduces exposure risk, but it sacrifices profitability. In contrast, only the pure time series model of LSTM was able to bring better forecasts and better buying and selling signals, bringing a higher profitability. The trade-off is between greater risk taking and higher profitability, greater portfolio control and lower profitability, which was proposed by the research. I understand that, in this scenario, we can also validate what the literature says about the more you take risk, the greater the return and the reverse path is also present in this experimentation.

And finally, analyzing the results of the models. The models had low performance metrics of quality, accuracy, and prediction. This demonstrates that it is dependent on data and its ability to predict is quite limited. The model where we used LSTM+HMM was slightly better performing. However, it demonstrates that perhaps a critical analysis, it is not the best model for this type of analysis and this type of data. This may put in check the initial purpose of our research. Because it has a certain complexity of temporal data processing and the result is what was evidenced in the performance metrics.

7 Conclusion and Future Work

This study sought to address the limitations of a static investor profile, proposing a dynamic in the disclosure of recommendations through a machine learning model. The approach chosen by the author sought to combine LSTM price forecasting with Markov HMM-Based volatile regime direction to provide buy and sell signals and portfolio reallocation for both institutional investors and ordinary investors. The author's proposal aimed to demonstrate that it is possible to develop a tool capable of interpreting beyond market data, market situational data, and being able to support decision making in portfolio reallocation. This application, if used, could be integrated into different brokers, investment analysis tools and decision support tools for the financial market in general.

The methodology implemented in a modular way, where the author sought to segregate the solution into different modules within a framework, where it starts by capturing market data through Yahoo Finance, goes through the data processing phase, where outliers are handled, and the normalization and equalization of these data for the model processing, the execution of the models itself, LSTM+HMM, HMM, and LSTM separately, and a last module responsible for making the visualization of these data as a whole. The data used in this methodology was based on two principles, a model of high volatility and market risk, which was the period of the pandemic, Covid-19, between 2020 and 2025, and a second sampling was made between 2012 and 2019, a period of greater market tranquility. The idea was to use assets with higher trading volume in the US financial market, such as ETFs SPY, QQQ, and shares of Apple (AAPL) and Microsoft (MSFT), for being a range of assets that give robustness by their breadth within the global investment landscape in general.

In the backtests, we used performance metrics of the portfolio, such as Sharpe Ratio, Maximum Drawdown, and the accumulated value, allowing comparisons on the effectiveness of the models and the effectiveness of the usability of the tool for the management of the investment portfolio of investors and institutions. The author sought to perform different experiments to demonstrate the use of the model and the proposed solution. The first involved the solution proposed in the initial work, LSTM+HMM, where it was possible to identify a return of adjusted risk and a portfolio growth in a perennial way over a period. In a second experiment we isolated the LSTM that presented a better performance from the portfolio point of view. In this experiment, it was possible to demonstrate that the HMM does not contribute significantly to the tested scenario and it may overlap the ability of the LSTM to capture volatilities. The HMM had a superior performance in buy-and-hold baselines, especially in periods of stability. This possibly demonstrates that the Deep Learning model, alone, seems sufficient to capture market change regimes, if trained with large windows.

Despite the promising results, it is fundamental to highlight some limitations in the analysis of both the model performance and portfolio performance that were proposed. The simulated portfolio uses only 4 assets, which gives a significant technical limitation. In addition, there is no use of macroeconomic data or technical indicators and exogenous variables that the authors could enrich the model as a whole. It is limited to an experiment on top of a hypothetical test, there is no portfolio bilaterality in this case. Another limiting point was the data processing capacity, because it does not have a significant infrastructure to process larger volume of data, there is also a limitation with regard to what was possible to train in this model. For future work and extensions of the research, we can compare different market regimes, enlarge the data collected in 5, 10, 20 year regimes to be able to observe what was the result of the models with this volume of data. The inclusion of new variables, such as interest rate, inflation, VIX index, etc. In addition to the application of more complex models, such as attention mechanisms (Attention), Transformers or Reinforcement Learning techniques, represents a promising path for portfolio allocation in more dynamic environments.

Finally, the author highlights that the research brings significant contributions to the financial and academic field. Highlighting that this research delivers a modular, replicable framework based on AI, focused on temporal decisions. Something that is still scarce in the available literature. There is a large volume of development being done, but little effectively worked. And this model, if expanded and replicated, can deliver significant market tools for both ordinary investors and institutional investors. Showing its performance and robustness as an investment asset management mechanism. Even with few assets, the model has already presented empirical evidence favorable to the dynamic time approach. Showing that not only the data purely on price, closing, and volume, but that market movements and external values influence the final result of an investment portfolio.

References

Ambuli, T. V., Venkatesan, S., Sampath, K., Devi, K. and Kumaran, S. (2024). Ai-driven financial management: Optimizing investment portfolios through machine learning, *2024 International Conference on Circuit Power and Computing Technologies (IC-CPCT)*, IEEE. Accessed: 01/02/2025.

Botta, A., Mohini, P. V., Raju, A. N. and Balaji, C. (2024). Deep learning for sentiment analysis in financial markets, *5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, IEEE. Accessed: 01/02/2025.

Daita, S. R., Choubey, S. B. and Choubey, A. (2022). A journal on cryptocurrency analysis and price prediction model using lstm neural networks, *2022 International Conference on Recent Trends in Microelectronics, Automation, Computing and Communications Systems (ICMACC)*, IEEE. Accessed: 01/03/2025.

Day, M.-Y. and Lin, J.-T. (2019). Artificial intelligence for etf market prediction and portfolio optimization, *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, IEEE/ACM. Accessed: 01/03/2025.

Dhuhita, A., Wibisono, A., Widodo, S. and Pambudi, M. (2023). Gold price forecasting using long short-term memory (lstm), *2023 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, IEEE. Accessed: 22/07/2025.

Dhuhita, W. M. P., Al Farid, M. F., Yaqin, A., Haryoko and Huda, A. A. (2023). Gold price prediction based on yahoo finance data using lstm algorithm, *2023 International Conference on Informatics, Multimedia, Cyber and Information Systems (ICIMCIS)*, IEEE. Accessed: 01/03/2025.

Ferdiansyah, Othman, S. H., Radzi, R. Z. R. M., Stiawan, D., Sazaki, Y. and Ependi, U. (2019). A lstm-method for bitcoin price prediction: A case study yahoo finance stock market, *2019 International Conference on Electrical Engineering and Computer Science (ICECOS)*, IEEE. Accessed: 28/02/2025.

Gill, K. S., Anand, V., Chauhan, R., Garg, A. and Gupta, R. (2023). Ethereum cryptocurrency prediction using ml procedures on recurrent neural network using lstm model, *2023 3rd International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)*, IEEE. Accessed: 01/03/2025.

Ilhan, F., Karaahmetoglu, O., Balaban, I. and Kozat, S. S. (2023). Markovian rnn: An adaptive time series prediction network with hmm-based switching for nonstationary environments, *IEEE Transactions on Neural Networks and Learning Systems* **34**(2): 715–728.

Johnson-Skinner, E., Liang, Y., Yu, N. and Morariu, A. (2021). A novel algorithmic trading strategy using hidden markov model for kalman filtering innovations, *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, IEEE. Accessed: 28/02/2025.

Komosar, T. and Kijanovic, V. (2023). Research on the application of open data in the public finance sector, *Economic and Social Development: Book of Proceedings*. ISSN: 1849-7535, Accessed: 22/07/2025.

Kour, M. (2024). Revolutionizing finance: Unleashing machine learning's potential for a new financial era, *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, IEEE. Accessed: 01/02/2025.

Kumar, S., Tuteja, A., Aggarwal, U., Matta, P., Gautam, P. and Maurya, S. (2023). Lstm approach for efficient stock market prediction, *2023 3rd International Conference on Intelligent Technologies (CONIT)*, IEEE. Accessed: 28/02/2025.

Pangavhane, P., Gadekar, T., Kolse, S., Darwante, N., Avhad, P. and Chaudhari, S. (2023). Transforming finance through automation using ai-driven personal finance advisors, *2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*, IEEE. Accessed: 07/02/2025.

Pradhan, A., Behera, S. and Sharma, P. (2023). Cnn-lstm hybrid model for high-frequency trading volume prediction, *International Conference on Data Engineering and Communication Technology (ICDECT)*, Springer. Accessed: 22/07/2025.

Preis, T., Moat, H. S. and Stanley, H. E. (2013). Quantifying the predictive power of browsing behavior on financial markets, *Scientific Reports* **3**: Article number: 1684. Accessed: 22/07/2025.

Reddy, R., Kumar, V. and Varma, M. (2024). Hybrid cnn-rnn model for financial fraud detection in banking systems, *International Journal of Advanced Computer Science and Applications (IJACSA)* . Accessed: 22/07/2025.

Seshagani, K., Raju, A. and Kavitha, R. (2025). Tcn-informer: A hybrid deep learning architecture for financial time series forecasting, *2025 International Conference on Data Science and Computational Intelligence*, Springer. Accessed: 22/07/2025.

Shahani, M. N. A., Rajapaksha, R. T. S., Maharoof, U., Nawinna, D., Senananda, V. M. and Attanayaka, B. (2023). Investopal - smart financial investment advisory system, *2023 7th International Conference on Computing, Communication, Control and Automation (ICCUBEA)*, IEEE. Accessed: 01/02/2025.

Shukla, N., Jaiswal, S., Chaudhri, C., Pinjarkar, L., Agrawal, P. and Kaur, G. (2024). An approach to the finance affair through reinforcement learning, *2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, IEEE. Accessed: 28/02/2025.

Varadarajan, M. N. and Priya, S. (2024). Ai and ml in finance: Revolutionizing the future of banking and investments, *2024 6th International Conference on Energy, Power and Environment (ICEPE)*, IEEE. Accessed: 01/02/2025.

Wei, Y., Tao, W., Zhang, H. and Kou, D. (2024). Research on cargo volume forecasting based on the arima-lstm hybrid model, *2024 3rd International Conference on Data Analytics, Computing and Artificial Intelligence (ICDACAI)*, IEEE. Accessed: 01/03/2025.

Wong, W., Ramachandran, P. and Pathak, S. (2023). High-frequency financial market prediction using machine learning: A case study on tesla, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)* . Accessed: 22/07/2025.

Xie, Z., Lin, X., Zhong, Y. and Chen, Q. (2019). Research on gold etf forecasting based on lstm, *2019 IEEE International Conference on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking (ISPA/BDCloud/SocialCom/SustainCom)*, IEEE. Accessed: 01/03/2025.

Yan, X., Chen, J., Zhao, L. and Zhang, Y. (2021). Research on the evaluation model of multi-source open data quality in the financial sector, *E3S Web of Conferences*, Vol. 275. Accessed: 22/07/2025.

Zaychenko, Y. and Kuzmenko, O. (2022). Investigation of hybrid deep learning networks and lstm in the short term forecasting in financial sphere, *2022 17th International Conference on Computer Sciences and Information Technologies (CSIT)*, IEEE. Accessed: 28/02/2025.