Using Statistical Learning Theory to Predict Gold Prices

Dissertation Project during master's program in Data Science - 2023

Mohammed Arslaan Kola

Abstract

Advanced analytical techniques are now a crucial component of effective decision-making in the realm of financial markets. The project at hand concerns itself with the convergence of statistical learning theory and technical analysis, aiming to predict market trends by utilizing a hybrid approach that incorporates Markov hidden models and Long Short-Term Memory (LSTM) networks. The primary objective of this undertaking is to construct a predictive model that deepens our understanding of financial market dynamics while achieving precise prognostic accuracy.

The amalgamation of Markov hidden models and LSTM networks presents a fresh outlook on predicting market trends. By associating the concealed states with varying market circumstances, the model apprehends intricate connections between indicators, providing valuable discernment into the fluctuations of financial markets. A significant amount of parameter tuning, and meticulous model selection is executed to ensure an optimal level of performance and dependability.

The model that was created exhibits its ability to forecast market trends and adjust to variable circumstances through practical examination. The outcomes of the assessment yield a more comprehensive comprehension of the correlation amidst monetary indications and market shifts. The significance of these discoveries stretches out to traders and investors, allowing them to make knowledgeable decisions and potentially enhance their trading tactics.

The present study adds to the growing body of research on financial market analysis by presenting a cohesive approach that merges technical analysis with statistical learning theory. Through leveraging Markov hidden models and LSTM networks, this investigation reveals their predictive value in comprehending and projecting market trends. The resulting forecasting model demonstrates potential in improving market trend predictions, thereby equipping traders and investors with an effective tool for navigating the intricate terrain of financial markets.

Table of Contents

List of Figures

List of Tables

1. Introduction

Finance and statistics have long had a strong association, as market participants attempt to detect patterns and trends in market data to influence their investment decisions. Recent technological breakthroughs and the emergence of machine learning have opened new avenues for financial market analysis. These advancements have resulted in the use of sophisticated algorithms and machine learning models capable of processing enormous amounts of financial data and uncovering hidden patterns and linkages. This incorporation of machine learning techniques has transformed financial analysis, allowing for more accurate forecasts and insights into market movements (Shen, 2020).

The application of machine learning algorithms in finance has enabled researchers and practitioners to improve market projections by using statistical analysis and pattern identification. These algorithms adapt to shifting market conditions and generate forecasts about future trends by learning from historical data. This has been useful in quantitative trading, risk management, and portfolio optimisation. With constant study and innovation stretching the frontiers of predictive skills in financial markets, the area of financial analysis is primed for additional improvements as technology continues to evolve. Finance, mathematics, statistics, and machine learning are defining the future of finance, giving traders and investors powerful tools to acquire a competitive edge in an increasingly complicated and dynamic landscape.

Technical analysis, which involves utilising statistical models to spot trends and potential trading opportunities in financial markets, is one topic of particular interest in financial analytics. Technical analysis primarily entails the examination of numerous financial market indicators, such as moving averages, Bollinger Bands, and the relative strength index (RSI), which are often employed by traders to identify probable buy and sell signals in financial markets.

This study aims to investigate the use of statistical learning theory in technical analysis by employing Markov hidden models to analyse three popular financial market indicators: Bollinger Bands, RSI, and EMA. The goal is to acquire insights into the underlying patterns and correlations between these indicators and to construct a predictive model that can be used to forecast market developments, addressing the challenges of identifying reliable signals and hidden factors influencing the behaviour of these indicators.

This project's justification originates from two basic motivations. To begin, there is a clear need in the field of technical analysis for a more complex and precise technique to projecting market movements. The goal is to build a more robust and effective strategy for anticipating market movements by applying statistical learning theory into the examination of financial market indicators. This has the potential to tremendously help traders and investors by providing them with more accurate and timely information that will allow them to make better informed investment decisions (Malato, 2020). The initiative aims to discover underlying patterns and relationships in data that can be used to improve the accuracy of market predictions, perhaps leading to higher profits and lower risk.

Second, this research intends to add to the growing body of knowledge in the field of statistical learning theory and its application to financial analysis. The research aims to reveal new insights and enhance the understanding of how financial market indicators interact and influence market movements by diving into the dynamics of financial markets and investigating the behaviour of financial market indicators. The project's goal is to push the boundaries of existing knowledge and contribute to the growth of the discipline by developing new models and methodologies for analysing financial markets. This research has the potential to improve the understanding of market dynamics and provide significant insights into the behaviour of financial market indicators, enhancing the field of financial analysis even more.

1.1 Problem Statement and Research Objectives

Financial markets are characterized by their intricate and dynamic nature, making accurate trend forecasting a critical challenge. Traditional approaches often fall short in capturing the complexities of market behaviour and fail to provide reliable predictions. To address this issue, the project aims to explore the application of statistical learning theory, particularly utilizing Markov hidden models, to analyse three widely used financial market indicators: Bollinger Bands, Relative Strength Index (RSI), and Exponential Moving Average (EMA). The overarching problem is the lack of accurate predictive models that can comprehensively forecast market trends and provide insights into the underlying dynamics of financial markets.

The research objectives guiding this study are as follows and the project timeline is depicted as a Gantt chart in Appendix 1:

- Data Collection for Financial Indicators: Collect historical data for Bollinger Bands, RSI, and EMA from credible financial data sources. Ensure that the collected data includes relevant market prices and variables necessary for accurate indicator calculation.
- Develop Markov Hidden Model: Construct a Markov hidden model to investigate the interplay between indicators and uncover hidden patterns. This model will capture the relationships and interactions between the indicators, revealing how they influence each other.
- Evaluate Model Performance: Assess the predictive model's success in forecasting future market trends using appropriate indicators such as accuracy, precision, recall, and the F1 score. Determine the model's capability to capture market movements and predict trends effectively.
- Compare with Existing Techniques: Compare the performance of the Markov hidden model with other commonly used statistical techniques in technical analysis, such as linear

regression and neural networks. This comparison will shed light on the model's strengths and weaknesses in relation to established methods.

- Investigate Sensitivity and Inaccuracy: Analyse the model's sensitivity to various parameters and hyperparameters. Investigate potential sources of inaccuracy or bias in the model's predictions, aiming to enhance its accuracy and reliability.
- Examine Practical Implications and Future Research: Explore the practical implications of the research findings for traders and investors. Outline potential future research directions in applying statistical learning theory to financial analysis, with the goal of enhancing prediction models and refining trading strategies.

1.2 Significance and Relevance of the Research

The research is highly significant in addressing the persistent challenges faced by traders and investors in accurately predicting market trends. By introducing a novel approach that combines Markov hidden models with statistical learning theory, this study aims to revolutionize trend forecasting. The potential of developing a predictive model that can more accurately anticipate market movements holds immense value for financial professionals seeking to optimize their decision-making processes.

The findings of this research could directly impact traders and investors by providing them with more reliable insights into market dynamics. Improved accuracy in forecasting can lead to better risk management, optimized investment strategies, and enhanced overall market performance. Furthermore, the research's focus on exploring hidden patterns and relationships among financial indicators can open new avenues for understanding market behaviour, potentially unveiling previously unnoticed trends and dependencies.

In conclusion, the significance of this research lies in its potential to bridge the gap between complex market dynamics and predictive modelling, offering practitioners a more advanced and dependable tool for navigating financial markets. Through the integration of modern statistical techniques and innovative approaches, this research aims to contribute to the advancement of financial analysis and prediction, ultimately benefiting traders, investors, and the broader financial community.

2. Literature Review

Forecasting stock prices is a formidable task due to the intricate and non-linear connections between various factors that influence stock prices. Traditional econometric models, such as time series models and regression analysis, have limitations in comprehending these complex relationships and patterns in the data, leading to imprecise predictions. Lately, researchers have focused on using machine learning methodologies to improve the accuracy of forecasting stock prices. For instance, Zhang et al., 2022 proposed a two-stage ARIMA model based on machine learning methods that leverages intraday transaction data of the stock market as auxiliary information to predict both linear and nonlinear features of the data. Empirical results manifest that this model effectively enhances prediction accuracy while being robust in predicting stock prices.

Ansah et al. (2022) have conducted a thorough analysis of various approaches to forecasting stock prices, highlighting the shortcomings of existing techniques and proposing alternate methods that can potentially lead to more precise and efficient predictions. In their research, they have identified Random Forest (RF), Long Short-Term Memory (LSTM), and Support Vector Machine (SVM) as some of the most widely adopted methodologies for accurately predicting stock prices. Bao and Yue's (2017) study have also corroborated this finding by demonstrating the effectiveness of these machine learning algorithms in capturing complex data patterns that can be used to make future projections with greater accuracy. It is evident from these studies that there is a need for more comprehensive techniques that take into account multiple factors when making such predictions, which may result in better outcomes for investors seeking reliable market insights.

In recent years, there has been promising progress with implementing machine learning algorithms to predict stock prices because they can capture complex patterns in large datasets, learn from past trends, as well as adapt to new developments - an essential feature for accurately predicting future values. To predict complex patterns in a dataset like predicting stocks accurately requires the use of decision trees, support vector machines (SVMs), neural networks or random forests; all are distinct types of machine learning approaches applied in forecasting stocks. Tsantekidis et al.'s (2017) research compared different algorithms' performance effectiveness for forecasting stocks - It was discovered that neural networks were most accurate at making predictions. Neural networks are especially effective when analysing large volumes of data, and they can identify complex relationships among different variables with ease.

2.1 Hidden Markov Model

Hidden Markov Models (HMMs) are mathematical tools that have found applications in finance for modelling time series data with complex dependencies. They are used when the underlying state of the system is not directly observable. HMMs consist of a set of states, transitions between these states, and observations generated based on the current state (Mor et al., 2020). In financial settings, HMMs have been instrumental in solving various problems, including portfolio optimization and stock price prediction.

For instance, by enhancing the predictive power of statistical models through HMMs, data scientists and algorithmic traders can improve their trading strategies (Skinner et al., 2021). Moreover, researchers have used HMMs to capture structural changes influencing prices and other economic indicators like institutional policies and regulatory authority intervention (Chang & Hu, 2022). By analysing market conditions using HMMs' transition law between different market states or regimes, it is possible to develop corresponding trading strategies that can yield profits via machine trading (Chen et al., 2020).

Thus, Hidden Markov Models provide a powerful framework for modelling complex financial data in situations where there are multiple hidden factors affecting asset prices or other economic indicators. Through their ability to capture structural changes and model hidden states or regime switches over time accurately, they offer insights into market conditions that human investors can use to make better investment decisions for optimal returns (Somani, 2014).

2.2 Technical Indicators

Technical indicators are mathematical calculations based on a stocks historical price and volume data that traders and analysts use to spot patterns and anticipate future price movements. Trends, momentum, volatility, and other features of a stock's behaviour can be identified using technical indicators (Agrawal et al., 2021). To make trading decisions, these indicators are frequently utilised in conjunction with other methodologies like as fundamental research and market sentiment. There are numerous forms of technical indicators, each with its own set of advantages and disadvantages. Moving averages, the relative strength index (RSI), Bollinger Bands, and the stochastic oscillator are among the most commonly used technical indicators.

In the world of financial analysis, experts often turn to the Exponential Moving Average (EMA) as a valuable technical indicator. Unlike its counterpart the Simple Moving Average (SMA), EMA provides more weightage to recent price fluctuations while calculating the average closing prices of stocks over a specific period (Tincher, 2022). This attribute makes EMA an ideal tool for spotting trends in the market and predicting future stock behaviour. Analysts also prefer to use EMA in conjunction with other technical indicators like Bollinger Bands and Relative Strength Index (RSI) for more accurate predictions and insights into market patterns. With its ability to respond quickly to changes in stock prices, EMA is regarded as an effective tool that can help traders make informed decisions about their investments.

The Relative Strength Index (RSI) serves as a highly valuable technical analysis tool for financial analysts to gauge the strength of a particular stock's price trend. Its widespread use is due to its effectiveness in facilitating the assessment of whether a stock has become overbought or oversold. The RSI relies on computing the average gains and losses for said stock over a stipulated period, offering investors an accurate understanding of where it stands within the 0 to 100 scale. If an RSI value surpasses 70, it indicates that the stock has become overbought, which could lead to an impending price decrease. Conversely, if its RSI value falls below 30, it suggests that the market has oversold this specific asset and it may be due for an upswing in its price (Fernando, 2023). These values play an integral role in assisting investors with their decision-making process by enabling them to decide whether they should hold onto their shares or sell them off.

The RSI reading provides traders with insight into how well-defined trends are within a particular market sector, which can be especially useful when determining entry and exit points for trading positions. As such, it remains one of the most popular technical indicators used amongst traders worldwide for both short-term and long-term investing strategies alike. Overall, the Relative Strength Index (RSI) can significantly enhance investor confidence by providing precise insights into market trends and movements of individual assets.

Bollinger Bands are an essential technical indicator in financial analysis, as they gauge a stock's volatility (Kumar & Archana, 2022). The bands are formed through the combination of a moving average (EMA) and two standard deviations above and below it. By analysing the width of the bands, traders can determine how volatile a particular stock is likely to be. A wider band indicates higher volatility compared to narrower ones that indicate lower volatility (Williams, 2013). It is important to note that Bollinger Bands usually work best when used in conjunction with other technical indicators like Relative Strength Index (RSI) and Exponential Moving Average (EMA). This allows for a more comprehensive analysis of the market trend, enabling traders to make informed decisions when it comes to buying or selling stocks that could potentially result in profits or losses.

In the realm of financial analysis, a range of diverse techniques are utilized to analyse and anticipate stock prices. The industry has identified Hidden Markov Model, Exponential Moving Average, Relative Strength Index, and Bollinger Bands as some of the most coveted tools for these purposes. Of these four techniques, Hidden Markov Models have gained significant recognition due to their ability to adeptly model intricate correlations in time series data. In contrast, technical analysts commonly rely on Exponential Moving Average (EMA), Relative Strength Index (RSI), and Bollinger Bands as widely used indicators for detecting trends in the market, gauging stock strength, and evaluating volatility levels. These tools serve as crucial components in the arsenal available to financial professionals when analysing stocks and making important investment decisions.

The utilization of diverse tools for predicting asset price fluctuations has significantly increased in recent years, indicating the effectiveness and reliability of these techniques. To enhance the accuracy of price trend forecasts, investors often incorporate machine learning algorithms or statistical models into their analysis. These predictive models help make more informed investment decisions, boosting profits and minimizing losses. In addition to traditional technical analysis methods, traders also leverage sentiment analysis of social media feeds to gather valuable data on market trends. This approach enables them to access additional insights from a vast array of sources while making informed trading decisions.

Overall, it is evident that combining various analytical tools with modern technology can lead to improved predictions about future financial market trends. By utilizing these techniques, investors and traders can stay ahead of the curve and achieve success in today's ever-evolving financial landscape. By gaining a comprehensive understanding of how to use different tools and techniques, investors can gain an advantage over their peers when it comes to making informed investment decisions. For instance, incorporating hidden Markov models (HMMs) with other methods such as technical indicators or sentiment analysis can result in more precise predictions about the future trends and movements in financial markets. By doing this, investors can stay ahead of market fluctuations and position themselves accordingly. Employing machine learning algorithms in conjunction with these tools can further enhance the accuracy of these forecasts and help investors make profitable investment choices. Thus, comprehending the combined impact of various approaches on investments is essential for successful investing.

2.3 Time Series Analysis

To make accurate predictions about stock prices, employing time series analysis techniques in conjunction with machine learning algorithms is the way to go. These techniques are specifically designed to model the time-dependent nature of data and, therefore, are adept at predicting stock prices. One commonly used time series model is the autoregressive integrated moving average (ARIMA) model which can capture a wide range of trends, seasonality, and cyclical patterns in stock prices.

A recent study by Gao (2021) compared different time series models including ARIMA to determine their accuracy in predicting stock prices. The findings demonstrated that ARIMA had the highest accuracy, making it an excellent choice for forecasting purposes. While machine learning algorithms have their advantage in uncovering complex relationships among variables, time series analysis techniques come into play when modelling behaviour that varies over time adding another layer of depth and complexity to data analysis. Therefore, a combination of both methods can be used by researchers to create comprehensive models that are more accurate and robust.

Depending on the problem at hand and nature of data being analysed, choosing between machine learning algorithms or time series analysis techniques will differ from project to project. Nevertheless, using both these approaches can prove beneficial as it helps consider multiple factors influencing stock prices.

2.4 Challenges

Stock price prediction is a challenging task as the data in financial markets is often noisy and unpredictable. One of the challenges is the presence of outliers and anomalies in the data, which can have a significant impact on the performance of machine learning algorithms and time series models. Afzal et al. (2021) proposed a method for outlier detection in stock price data using a combination of statistical methods and machine learning algorithms. Their proposed method involves pre-processing the data, extracting features, applying statistical methods to identify outliers, and then using machine learning algorithms to remove them. The method was evaluated on several datasets and was found to be effective in detecting and removing outliers, thereby improving the performance of machine learning algorithms in stock price prediction.

Another challenge of stock price prediction is the interpretability of machine learning models. Machine learning algorithms are often considered black boxes, which means that it is difficult to understand how they arrive at their predictions. Interpretability of machine learning models is also an important issue in the stock price prediction. The interpretability of machine learning models means understanding how the model makes predictions or decisions. This is important in finance since investors need to have confidence in the models that they use for making investment decisions. Yun, Yoon, and Won (2022) proposed a method for interpreting the results of machine learning models for stock price prediction by identifying the most important features in the data. The method is based on permutation feature importance, which is a technique for measuring the importance of input features by randomly permuting their values and observing the effect on the model's output.

2.5 Bid-Ask Price

Financial markets can be a daunting landscape to navigate as they are multifaceted and possess several intricacies that are necessary to understand in order to make sound investment decisions. One such feature is the bid-ask spread, which indicates the divergence between the price buyers are willing to pay for an asset compared to what sellers will accept. However, acquiring accurate data on these prices can prove challenging due to various factors. One of the primary challenges of obtaining accurate data on bid-ask spreads is market fragmentation. Markets may have multiple trading venues with no centralized data, which could create pricing discrepancies across different platforms. This makes it difficult for traders to obtain consistent and comprehensive information about bid and ask prices.

Another challenge arises from delays in transmitting and processing data, making it challenging for traders to access up-to-date information on bid-ask spreads. In addition, high-frequency trading (HFT) - a trading strategy using algorithms at lightning-fast speeds - poses a significant challenge as it creates rapid fluctuations in bid and ask prices that traditional data sources struggle to track accurately. Understanding how market participants behave is another critical factor that influences bid-ask prices. The state of the market also plays a significant role in determining these prices, further complicating matters having intricate interplays between many variables other than just the spread itself.

In conclusion, financial markets' complexities mean there is more to consider beyond just analysing bid-ask spreads. Successful traders need a keen understanding of various factors such as market fragmentation, HFT strategies, behavioural patterns of participants all while considering broader economic conditions influencing price movements if they hope to make informed investment decisions.

2.6 Risk Assessment

- Data Quality and Accuracy: In the world of stock market trading, accurate and high-quality data is an integral component for successful stock price prediction. The slightest errors, missing information, or inconsistencies in the data can significantly impact the accuracy of predictions (Mahanti, 2022). As such, it is imperative to ensure that any data utilized in the process of predicting stock prices comes from trustworthy sources and is regularly updated to maintain its reliability and suitability for use. This way, traders can make informed decisions based on timely and precise information concerning stocks' prices and trends. Thus, it is crucial to pay close attention to the accuracy and quality of data used in making any predictions related to stock trading.

- Model Performance: It is important to understand that when using machine learning algorithms or time series analysis techniques, there is a possibility that the models selected may not provide the desired results as anticipated. This can be due to various factors such as these models being unable to identify the intricate patterns and relationships present in complex stock market data, potentially leading to incorrect forecasts. To prevent this from happening, it is vital to conduct thorough testing and validation of these models before application. This process is essential for both identifying and minimizing any potential risks associated with using machine learning algorithms or time series analysis techniques in predicting stock market trends (Pustokhin & Pustokhina, 2022).

- Overfitting and Generalization: Overfitting is a common issue that arises in machine learning models where the model becomes overly complex and starts to fit too closely to the training data. This can lead to inaccurate predictions on new or unseen data, which can ultimately hinder the performance of the model (Ying, 2019). Therefore, it is important to take measures such as cross-validation and regularization techniques to prevent overfitting. Cross-validation helps in evaluating the performance of a model by splitting up the data into training and testing sets. By doing so, it enables us to check how well the model performs on unseen data while also ensuring that we are not overfitting the model to specific training examples. Regularization is another way of reducing overfitting by adding constraints or penalties to limit the complexity of a machine learning model. This ensures that even if a particular pattern exists only in the training set, it will not be captured at an incredibly high degree leading to poorer generalization performance (Corsaro et al., 2022). Preventing overfitting is pivotal for achieving reliable predictions. Therefore, it's essential that machine learning professionals remain vigilant about this problem throughout their modelling process and use appropriate techniques such as cross-validation and regularization when necessary.

- Market Volatility and Uncertainty: The stock market is a complex system that is susceptible to numerous external factors, including fluctuations in economic conditions, geopolitical events, and regulatory changes. These factors can have a significant impact on the performance of stocks and can cause sudden volatility or unexpected price changes. This unpredictability makes it challenging to forecast stock prices with accuracy (Qiu et al., 2020). To mitigate the risk associated with these unpredictable shifts, it is essential for investors to adopt strategies that help them diversify their investments and incorporate multiple data sources when analysing the market. By spreading investments across different sectors and asset types, investors can reduce their exposure to any one particular risk factor (Söderblom et al., 2016). Additionally, utilizing various sources of data such as financial statements, news articles, and expert opinions can provide a more comprehensive understanding of the broader market trends that impact specific companies or industries. While external factors will always play a role in shaping stock market trends, investors who take proactive steps to manage risk through diversification and data analysis are better positioned to make informed investment decisions and achieve long-term success in this dynamic marketplace.

- Computational Resources and Infrastructure: The effectiveness of stock price prediction models is largely dependent on the complexity and size of the data involved. Some models may require massive computational resources and robust infrastructure to operate optimally, given their complex algorithmic processes. In instances where there are inadequate resources, it can result in delays in processing times, sluggish model training that slows down decision-making processes, and difficulties encountered when handling extensive datasets (Xu & Yang, 2023). Therefore, it's essential to have adequate computing power and efficient infrastructure in place to ensure seamless operations and timely delivery of insights for effective decision-making.

- Regulatory and Legal Risks: Complying with regulatory requirements and legal considerations is of paramount importance in the field of finance. This is because adherence to relevant regulations and laws such as data privacy and financial regulations can spare an organization from costly legal entanglements that could otherwise impede its progress (Allen & Koshima, 2018). Therefore, it's critical to stay informed on any changes made to the regulatory landscape and ensure that the project adheres strictly to all necessary guidelines. Failure to do so could result in severe consequences for both the organization as well as its customers or stakeholders. It is, therefore, crucial to take these matters seriously when handling finances – both for personal purposes and managing a business venture.

- Model Interpretability and Explainability: One significant challenge in using machine learning algorithms is the lack of interpretability associated with their predictions. This presents a potential risk, particularly for investors and stakeholders who may need clear explanations and justifications for the models' outputs. Without a clear understanding of how the model arrived at its conclusions, it could be harder to build trust among users or address any misinformation generated by faulty predictions. To manage these risks, techniques that enhance the interpretability and explainability of these models should be carefully considered and implemented. These techniques may involve providing visualizations, feature importance scores, or detailed documentation that explain how specific variables contribute to the final prediction. By implementing such measures, we enable users to gain more confidence in the models' outputs while also demonstrating transparency around how they operate.

- Financial Loss: Investing in the stock market comes with its own set of challenges and risks. One of the primary concerns is predicting stock prices accurately. Inaccurate or unreliable predictions can lead to substantial financial losses for investors or stakeholders who rely on such predictions to make investment decisions (Ghasemi & Ghasemi, 2020). Therefore, it becomes imperative to develop and implement proper risk management strategies to mitigate these risks.

One way to minimize potential losses is by diversifying investments across multiple sectors and industries, which helps reduce the impact of any one company's performance on overall returns. Another strategy is incorporating risk mitigation techniques such as stop-loss orders, which function as an automatic trigger that sells stocks once they hit a specific price limit (McNulty, 2022)

It's crucial that investors understand that there are no guarantees when it comes to investing in stocks, and even expert analyses can be wrong from time to time. However, by adopting effective risk management measures like diversification and risk mitigation techniques, investors can better manage their investments and minimize potential financial losses in situations where predictions don't pan out as expected.

- Ethical Considerations: In recent times, the practice of utilizing sentiment analysis or social media data for predicting stock prices has become increasingly popular. However, this trend is not without its fair share of ethical concerns. There are issues relating to data privacy and user consent that must be taken very seriously by all parties involved in such practices. Furthermore, there is also the concern regarding potential biases that can arise from the use of such data sets in stock market predictions. To ensure that these ethical concerns are properly addressed and managed effectively, it is essential that all researchers and stakeholders involved in this area take proactive steps towards handling and analysing such data ethically. Additionally, compliance with applicable regulations as well as adherence to established guidelines must be a top priority at all times. Only then can we ensure that these technologies are used responsibly for social good rather than being exploited purely for financial gain at the expense of individuals' rights and privacy.

- Competitive Landscape: In the domain of forecasting stock prices, it is a cut-throat competition where many players are involved in the race for dominance. These players can be financial institutions, research and analysis companies or simply individual traders who are always on the lookout for better ways to outdo their rivals in terms of accuracy, speed, and innovation (Sayadi, 2013). Falling behind in any of these aspects may lead to an irreparable damage to one's reputation and even loss of business. Therefore, it is highly recommended that one should keep a close watch on the competitive landscape by constantly monitoring all new advancements and keeping oneself abreast with the latest trends in this field. This helps businesses stay ahead of their competitors by making informed decisions based on up-to-date information.

In order to minimize the multitude of potential risks associated with stock price prediction projects, it is crucial to develop and implement proper risk assessment procedures, as well as effective mitigation strategies. Such measures can serve to safeguard against unexpected market turbulence and unforeseen events that could significantly impact the accuracy and reliability of predictions. Additionally, continuous monitoring is necessary to ensure that any emerging issues are identified and addressed in a timely manner.

Regular evaluation of models and methodologies is an integral part of successful stock trading. As market conditions tend to change every now and then, it becomes crucial to keep testing the effectiveness of existing models. Moreover, new research findings may emerge which could call for a shift in the approach used in predicting stock prices. To ensure that investors are not left behind due to reliance on outdated models, constant review and evaluation are essential.

These evaluations pave the way for adaptations that serve to improve the overall accuracy and reliability of predictions made regarding stock prices. Accuracy is key when it comes to investment decisions since inaccurate or unreliable forecasts can result in significant losses. Therefore, by making use of these evaluations, investors can make more informed decisions based on improved prediction models while minimizing potential losses associated with outdated or unreliable forecasts.

2.7 Transfer Learning

Transfer learning involves training a model on a source task and then using the acquired knowledge to improve performance on a target task. In the context of stock price prediction, transfer learning can be applied by training a model on related financial data or a similar stock before fine-tuning it on the target stock. This allows the model to leverage patterns and relationships learned from the source task to enhance its predictions on the target stock, even when there is limited training data available. The study conducted by Nti, Adekoya, and Weyori in 2020 highlights the effectiveness of transfer learning in improving the performance of machine learning models for stock price prediction.

Sentiment analysis is an important technique that involves examining the emotional tone or attitude conveyed in various sources of textual data including news articles, posts on social media platforms, and other similar sources. This practice can help analysts gain valuable insights into investor behaviour and market trends which can be useful for predicting stock prices. A recent study conducted by Jin et al. in 2020 specifically examined the impact of sentiment analysis conducted on news articles on the accuracy of stock price predictions. The researchers found that by considering the sentiment conveyed in news articles related to a particular stock or the overall market, it became easier to understand the prevailing market sentiment and potentially anticipate future price movements with more precision than ever before. This demonstrates how important sentiment analysis can be when making informed investment decisions based on predictive analytics.

By combining transfer learning with sentiment analysis, one can further enhance the predictive capabilities of stock price models. For example, a model pretrained on financial data from multiple stocks can capture general market trends and patterns (Yang et al., 2022). Then, by incorporating sentiment analysis of news articles or social media data specific to the target stock, the model can adapt its predictions to reflect the sentiment-driven dynamics of that stock. This integration of transfer learning and sentiment analysis provides a comprehensive approach to stock price prediction, leveraging both historical financial data and the current market sentiment.

While transfer learning and sentiment analysis can certainly improve the accuracy of stock price predictions, it's important to keep in mind that they are not infallible. The financial markets are incredibly intricate and unpredictable, influenced by many complex factors such as economic conditions, the performance of individual companies, geopolitical events, and many other variables. Even though these techniques can be helpful in enhancing prediction accuracy to some extent, they cannot guarantee precise forecasts due to the nature of their limitations. It's essential to consider both transfer learning and sentiment analysis as complementary tools that should be used alongside other approaches for predicting stock market behaviour. Ultimately, relying solely on these methods could lead to disastrous consequences since there is always a degree of uncertainty when it comes to predicting stock prices accurately.

3. Methodology

This section provides a detailed overview of the study approach used to investigate the application of statistical learning theory in technical analysis, with a focus on analysing financial indicators using Markov hidden models and LSTM networks. Data gathering, pre-processing, and the use of both Markov hidden models and LSTM networks are all part of the methodology.

3.1 Data Collection and Pre-processing

- Historical Data Source: In order to undertake the analysis, historical data on gold from Dukascopy, 2023 was acquired. The dataset consists of hourly gold prices denominated in USD and spans from January 2004 to July 2023. It features a range of key metrics such as opening and closing prices, high and low prices, trade volume, and timestamps. Table 1 shows the first five rows of the dataset. This extensive dataset forms the fundamental basis of the research by providing us with the means to scrutinize and explore the fluctuations in gold prices over an extensive period.

Gmt time	Open	High	Low	Close	Volume
01.01.2004 00:00:00.000	414.922	414.922	414.386	414.596	40970.0014
01.01.2004 01:00:00.000	414.887	414.89	414.391	414.391	25879.9996
01.01.2004 02:00:00.000	414.887	414.89	414.25	414.617	36159.9997
01.01.2004 03:00:00.000	414.887	414.89	414.474	414.89	19819.9991

Table 1: First 5 rows of the data

- Data Pre-processing: Prior to implementing any modelling techniques, the initial step taken was the pre-processing of data to guarantee its reliability and appropriateness for analysis. The process involved addressing missing values by replacing them with a value of zero and transforming the 'Gmt time' column into a datetime format for accurate temporal analysis.

- Feature Engineering: To bolster the dataset's representational prowess, fundamental metrics for technical analysis were derived. Esteemed financial analysis libraries were employed to calculate benchmarks such as the Relative Strength Index (RSI), Exponential Moving Average (EMA), and Bollinger Bands. Appendix 2, provides the Python code used to calculate the RSI, EMA, and Bollinger Bands using the talib library. Each indicator serves a specific purpose in technical analysis, aiding in understanding price trends, momentum, and potential reversal points. The calculated indicator values have been added to the Data frame for further analysis. Figure 1 shows the implementation of gold prices using candlesticks chart.

Figure 1: Candlestick plot for gold prices

3.2 Markov Hidden Models for Financial Indicators Analysis

- Introduction to Markov Hidden Models: The application of Markov Hidden Models, alternatively known as Hidden Markov Models (HMMs), aims to capture the underlying states that generate observed sequences through statistical modelling. By utilizing HMMs in this project, we intend to analyse the interaction among financial indicators, with the objective of revealing concealed patterns and interdependencies (Popov et al, 2021). This form of statistical modelling is a valuable tool for gaining insights into complex financial systems.

- Mapping Hidden States: In this project, the following concealed states, including Oversold, Overbought, Bullish, and Bearish, were distinguished by considering diverse factors associated with the indicators. These factors encapsulate distinct market situations, including optimistic trends, pessimistic trends, horizontal markets, and others. Every datum was allocated to a particular concealed state based on these circumstances. Appendix 4 presents the Python code for analysing the data and classifying hidden states based on the defined conditions and technical indicators. By categorizing data points into different states, we gain insights into the potential trends and conditions in the gold market. The count of data points in each hidden state helps us understand the distribution of states in the dataset.

- Quantization of Observations: To make it easier to use Hidden Markov Models (HMMs), a technique was employed where continuous data observations were transformed into discrete symbols. This was achieved by quantizing the 'Close' price and putting it into multiple bins, effectively converting it into categorical data. Additionally, other relevant features were also treated this way to better fit them into the model. The process of quantization allowed for more efficient and effective analysis of the data, making it easier to spot patterns and trends that might be difficult to discern from raw continuous data (Tadic & Doucet, 2003). Therefore, this approach proved useful in facilitating the application of HMMs and improving overall accuracy.

- HMM Parameter Estimation: The Gaussian Hidden Markov Model technique was implemented to effectively scrutinize and decipher the quantized observations and their respective covert states. To achieve this, parameters of the model such as transition probabilities and emission probabilities were estimated accurately with the aid of the Baum-Welch algorithm - an Expectation-Maximization technique. The Baum-Welch algorithm proved to be highly effective in determining these significant values for the model's parameters, which ultimately improved its overall efficiency in analysing the complex data set.

3.3 LSTM Networks for Time Series Analysis

- Introduction to LSTM Networks: Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) are distinguished for their capacity to detect long-term associations in sequential data (Bahad et al., 2020). The aim of this undertaking was to exploit LSTMs for projecting financial market trends by examining past data.

- Data Sequencing: To prepare the dataset for training and evaluating the LSTM network, it was divided into smaller segments of a specific length, forming input-output pairs. This segmentation allowed for easier processing of the data and helped in improving the accuracy of predictions made by the network. The sequences were composed of various features that were found to be relevant to predicting stock prices, paired with the corresponding 'Close' price that followed afterward. By feeding these input-output pairs to the LSTM model, it could learn patterns within each sequence and make informed decisions about future values based on past observations.

- Network Architecture: Utilizing bidirectional LSTM layers in the network architecture of LSTM has been found to be effective in capturing temporal dependencies from both directions, thereby augmenting the predictive capacity of the model. Additionally, attention layers were incorporated to weigh the significance of distinct time steps within the sequence.

- Loss Functions and Optimizers: The compilation of the model was accomplished through a fusion of loss functions. In order to predict the 'Close' price, the loss function employed was mean squared error (MSE). On the other hand, categorical cross-entropy loss was utilized for prediction of hidden state. The chosen optimizer for training was rmsprop.

3.4 Results Analysis

Following the completion of training for both the Markov hidden model and the LSTM network, a comprehensive evaluation of their respective outcomes was carried out. This evaluation involved an examination of numerous performance metrics designed to gauge their effectiveness in accurately modelling the data. To obtain a thorough understanding of how each model operates, it is essential to consider its performance from various angles. As such, a variety of metrics were deployed, encompassing aspects such as accuracy, sensitivity, and specificity. By analysing these metrics in-depth, it becomes possible to determine which model is most effective for the given task at hand.

The integration of Markov hidden models and LSTM networks in a comprehensive approach is aimed at acquiring an extensive comprehension of the fundamental dynamics characterizing financial markets. This integrated approach is expected to enhance the precision of forecasting trends in financial markets. The use of this amalgamated methodology provides insights into obscured patterns, which consequently presents valuable predictive capabilities, offering considerable contributions in improving trading strategies and investment decision-making.

4. Financial Market Indicators

In the realm of finance, the dynamics of financial markets are intricate due to the interplay of various factors and influences that determine asset prices (Chen et al., 2021). To make informed decisions amidst these constantly changing conditions, traders and investors employ diverse strategies and resources for analysing and projecting market behaviours. Among such resources, financial market indicators hold substantial significance in offering valuable observations on market trends, momentum, fluctuations, and potential reversals. These indicators are numerical metrics obtained from past price and volume data with the objective of identifying regularities and inclinations that could assist in decision-making.

Financial market indicators play a crucial role in technical analysis, a field that primarily examines past market data to predict future price fluctuations. These indicators enable traders to quantify and analyse various aspects of market behaviour, furnishing them with useful information to forecast trends and make informed trading decisions. The classification of financial market indicators is broad but can generally be categorized into trend-following indicators, oscillators, volatility indicators, and volume-based indicators.

One of the most effective ways for traders to get a sense of what's happening in the markets is to use trend-following indicators. These indicators help traders discern which direction prices are moving - if they're trending upwards (bullish) or downwards (bearish). It's important to know whether an asset is bullish or bearish because that can have a big impact on your trading strategy. Another type of indicator that traders often use is oscillators. These measure the momentum and overbought or oversold conditions of an asset, allowing traders to pinpoint potential reversals (Ni et al., 2020). This can be critical information for traders because catching a trend reversal early can be very profitable.

Volatility indicators are another tool in a trader's arsenal. They help quantify the degree of price fluctuations in a market, which in turn helps assess potential risks and determine how much exposure to take on any given position. Finally, volume-based indicators provide valuable insight into the strength or weakness of price movements based on trading volume. Trading Strategy guides (2023) mentions that by analysing trading volume alongside price action, traders can develop a more nuanced understanding of what drives market movements and how best to capitalize on them.

The present report investigates the comprehension and implementation of three significant financial market indicators: Bollinger Bands, Relative Strength Index (RSI), and Exponential Moving Average (EMA). These indicators offer distinct viewpoints on market analysis, empowering traders to evaluate market sentiment, recognize probable entry and exit points, and manage risks efficiently. By examining their fundamental principles, computations, and interpretive subtleties, the goal is to furnish readers with a comprehensive perception of how these indicators enrich technical analysis and improve the decision-making process in financial markets.

4.1 Bollinger Bands

Bollinger Bands are an extensively employed tool in technical analysis, serving the purpose of comprehending the potential price movement and volatility of a financial asset. These bands comprise three lines that are portrayed around the asset's price chart, with the centre line typically denoting an Exponential Moving Average (EMA), whereas the upper and lower bands are calculated based on standard deviation of the asset's cost. The importance of Bollinger Bands lies in their ability to demonstrate market volatility visually (Seshu et al., 2022). The bands expand when there is greater volatility, whereas they contract when it declines. Traders frequently utilize Bollinger Bands to identify probable price reversals, overbought or oversold conditions, along with deciphering the general range within which prices are expected to fluctuate. Figure 2 shows the implementation of Bollinger bands on the gold chart.

To calculate and interpret Bollinger Bands, there are various steps that traders must follow. The first step is to derive the centre line by using the Exponential Moving Average (EMA) which smoothens out any short-term fluctuations in asset prices. Once this is done, the upper and lower bands are then set at a specific number of standard deviations away from the centre line.

Figure 2: Bollinger Bands on Gold Data

The upper and lower bands then represent potential resistance and support levels respectively, which traders can use to anticipate future price movements. By analysing how the asset's price interacts with these bands, traders can get an idea of where it may be headed next. However, there are certain challenges and limitations associated with using Bollinger Bands as well. For instance, false signals can occur in highly volatile markets due to sudden price movements that cause the asset's price trend to diverge from what was previously expected based on historical data. Additionally, abrupt market shifts can also cause unexpected price trends that may not align with what is predicted by Bollinger Bands alone. As a result, it's important for traders to use other indicators alongside Bollinger Bands to make more informed trading decisions based on multiple factors.

4.2 Relative Strength Index (RSI)

The Relative Strength Index (RSI) is a technical analysis tool that is widely used to evaluate the speed and change of price movements. It works as a momentum oscillator, and it is considered an essential instrument for traders to understand overbought or oversold conditions in the market. By comparing recent gains and losses over a specific time period, RSI computes the strength of a price trend. Day et al., 2023 indicates that the RSI value ranges from 0 to 100 with readings above 70 often indicating overbought conditions and readings below 30 suggesting oversold situations. Due to its ability to signal potential trend reversals and confirm price trends' strength, The Relative Strength Index (RSI) has become an indispensable tool for investors in stock trading, forex markets, and other securities markets.

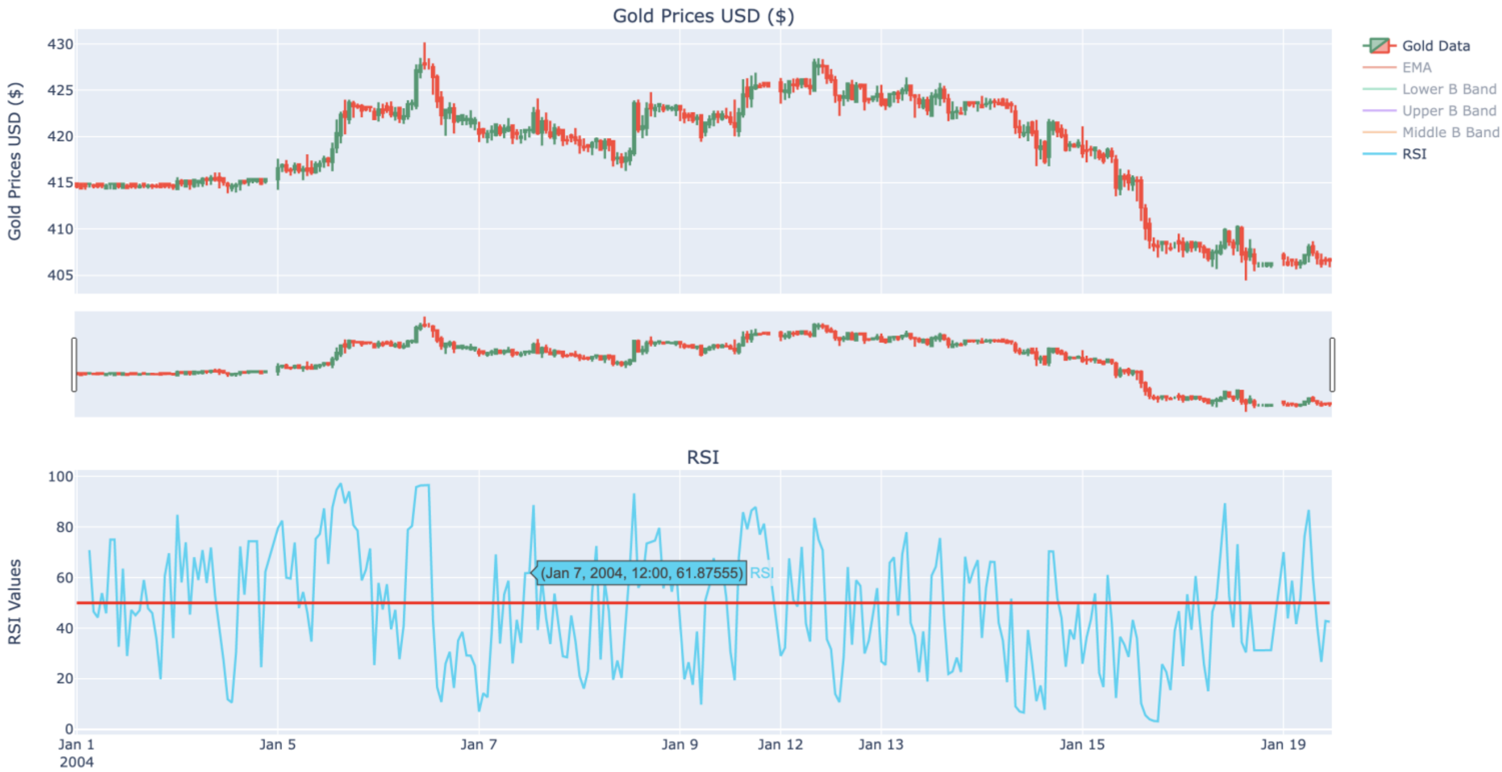

Figure 3: RSI on Gold Data

Figure 3 depicts the RSI chart below the gold prices. Comprehending RSI necessitates a grasp of its thresholds. If the RSI exceeds 70, it could imply that the asset is overbought, thus pointing towards an impending correction or reversal. Conversely, if the RSI falls below 30, it may indicate that the asset is oversold and suggest a possible price rebound. One of RSI's positive aspects encompasses its capability to offer early indications of trend reversals while being simple to use. Nonetheless, some shortcomings may also surface, including generating false signals during robust trends and its susceptibility to the selected time period. Traders must take these strengths and weaknesses into account when employing RSI as part of their analysis.

4.3 Exponential Moving Average (EMA)

The Exponential Moving Average (EMA) is a variant of moving average that grants more significance to current prices, thereby smoothing the depiction of pricing patterns. The EMA's value stems from its potential for offering traders insights into the direction of the current trend, making it an advantageous tool for identifying promising market entry and exit points. As compared to simple moving averages, the EMA reacts faster to changes in price due to its emphasis on recent data.

The Exponential Moving Average (EMA) is a financial tool that involves the application of a smoothing factor to the preceding EMA and adjusting it based on current market trends. This method of calculation enhances the EMA's responsiveness to recent price fluctuations, making it a valuable technique for traders seeking to identify early shifts in trend directions. According to Scott et al., 2016, one important aspect of using EMA in technical analysis involves contrasting the prevailing market price with the EMA and analysing crossovers between different periods of EMA as an indicator of potential changes in market trends. However, sudden spikes or rapid changes in prices can adversely impact the accuracy and effectiveness of EMAs, especially during volatile market conditions. Therefore, traders should consider both its advantages and limitations when incorporating this tool into their analytical strategies. Figure 4 shows the EMA line on the gold chart.

Figure 4: EMA on Gold Data

5. Model Development

This section provides an in-depth understanding and exploration of the complex procedure involved in creating predictive models. These models utilize statistical learning theory, which involves advanced techniques that aid in forecasting market trends accurately. The approach to be discussed combines two powerful methodologies, namely Markov Hidden Models and Long Short-Term Memory networks. Markov Hidden Models (HMM) are known for their sophistication when it comes to modelling sequences or time-series data effectively. On the other hand, Long Short-Term Memory (LSTM) networks serve as an extension to recurrent neural networks and are very efficient at processing sequential data with long-term dependencies.

The combination of these two approaches creates a synergistic framework that amplifies their strengths while offsetting each other's weaknesses. This results in an incredibly powerful predictive model capable of providing enhanced market trend forecasting capabilities. By delving deeper into this process, you can gain a better appreciation for the intricate details involved in developing such models. In turn, this knowledge can help you create more accurate forecasts that enable you to make better-informed decisions concerning your business or investments.

5.1 Hidden Markov Models (HMM)

The initiation of forecasting algorithm begins with the adoption of Hidden Markov Models (HMM). As per the research conducted by Chang & Hu, 2022, HMM is a statistical tool that captures the inherent dynamics of sequences of observations from various financial market indicators. This analysis allows us to understand and interpret how different variables interact within the financial markets. The core aspect of HMM is its ability to represent the association between hidden states and observable data, which assists in inferring potential underlying trends or behaviours that affect market patterns. These states remain unobservable but provide insight into understanding latent factors involved in financial market movements. Hence, HMM proves to be an effective approach for capturing these obscure variables within financial data.

The architecture of the HMM is created with utmost care and attention to detail by hand-picking the right number of hidden states and efficiently initializing the model's parameters. The next step involves training the model with the use of historical data from financial markets, where it can develop a deep understanding of patterns and correlations between observable indicators and underlying hidden states. The HMM is essentially a pattern recognition tool that helps identify repetitive behaviours that significantly impact market fluctuations, leading to profitable investments or informed decisions about buying/selling assets (Chen et al., 2020). With its ability to analyses vast amounts of information quickly, an HMM has proven to be a powerful tool in today's ever-changing financial landscape. Figure 5 shows the detection of two market states, Bullish and Bearish using Gaussian Hidden Markov Model. Appendix 3, provides the Python code that calculates the returns of the gold data and fits a Hidden Markov Model with bullish and bearish states. This model can be used to gain insights into the underlying market regime shifts and potential changes in price trends based on the calculated returns. The model's predictions, represented as probabilities of the hidden states, provide a framework for understanding market dynamics.

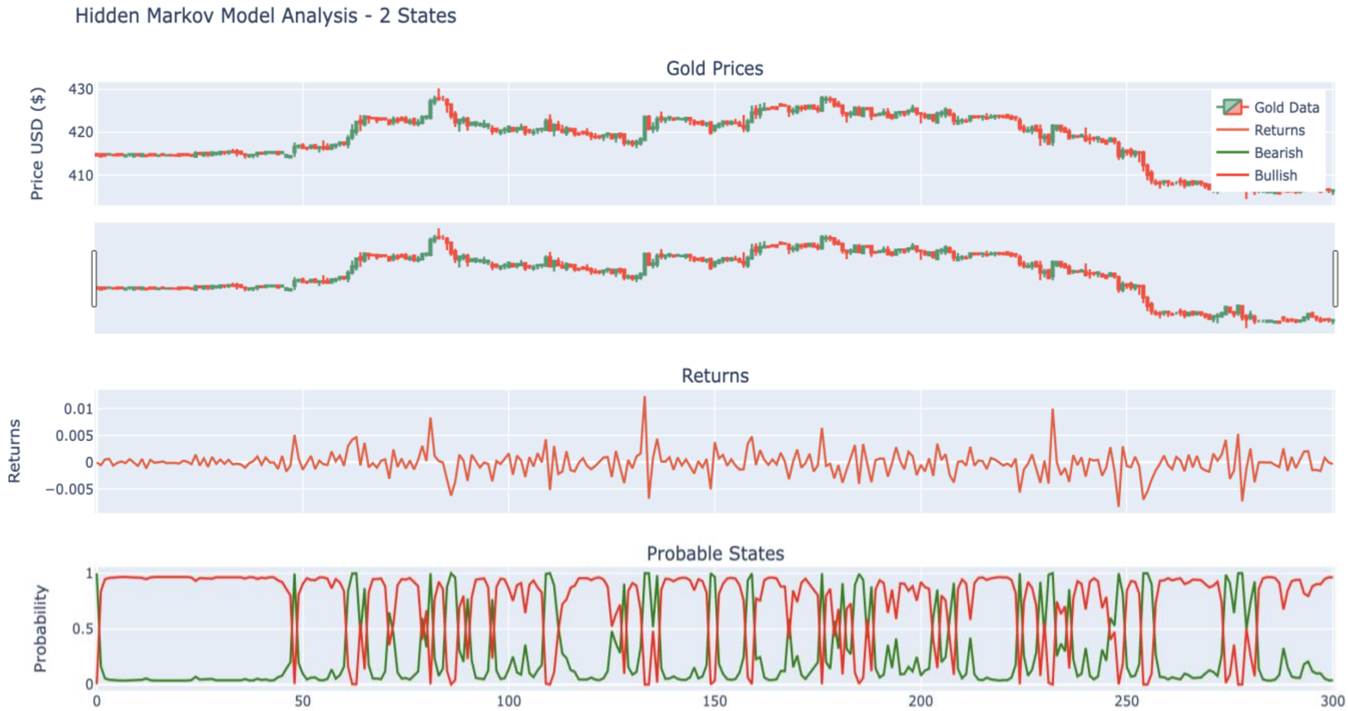

Figure 5: State detection using HMM

5.2 Integration of Markov Hidden Models and LSTM

To improve the precision and durability of a predictive model, it is crucial to blend the valuable insights obtained from HMM with the exceptional power of Long Short-Term Memory (LSTM) networks. LSTM has been recognized as an exceptional type of Recurrent Neural Network (RNN) that excels in detecting sequential dependencies and patterns embedded in time-series data sets. By combining HMM and LSTM, the design can be a hybrid structure that leverages both statistical modelling and deep learning capabilities, thus enhancing the overall performance of predictive models. Such a merger can take advantage of each method's unique strengths while minimizing their weaknesses, resulting in more accurate predictions with increased robustness.

The process of integration entails utilizing the results generated by the HMM as inputs for the LSTM. The said outputs signify the deduced covert states and offer a contextual basis for the LSTM to conduct deeper analysis and anticipate market inclinations. The memory cells of the LSTM hold crucial details about preceding sequences of covert states, granting it the ability to comprehend intricate temporal associations within the data. By fusing these techniques, a more comprehensive comprehension of market dynamics is achieved, leading to an enhanced capacity for forecasting trends.

5.3 Parameter Tuning and Model Selection Considerations

The development of an effective predictive model requires careful attention to parameter tuning and model selection. The latter involves optimizing various aspects of the HMM and LSTM, including but not limited to the number of hidden states, learning rate, and network architecture. This crucial step significantly enhances the model's performance while ensuring its alignment with specific characteristics of financial data.

When it comes to selecting the best model for market trend forecasting, there's more to consider than just adjusting parameters. One must also consider which evaluation metrics are most suitable for the task at hand. Metrics such as accuracy, precision, recall, F1 score, and others can help determine how well a model is performing in accurately predicting future trends. It is essential to conduct both thorough validation and cross-validation to test the model against various market scenarios within different periods of time. By doing so, it can be ensured that chosen model will be robust enough to withstand potential changes or fluctuations in the market conditions that may occur over time.

6. Results and Analysis

6.1 Presentation of empirical findings and outcomes

The following section entails the results obtained from the project that aimed to predict gold prices. The method for achieving this goal involved implementing an amalgamation of various statistical learning techniques, technical indicators, and advanced machine learning models such as Hidden Markov Models (HMM) and Long Short-Term Memory (LSTM) neural networks. To ensure accuracy and reliability in the outcomes, a meticulous approach is adopted that encompassed several stages. Initially, the data was pre-processed by cleaning it up and organizing it into manageable formats. Next, specific technical indicators are calculated to gold prices using various mathematical formulas. Afterward, the data is visualized to identify trends quickly and facilitate better understanding of market behaviour. Then the data is proceeded to train the models with immense attention paid to detail during every stage of model training to achieve optimal performance from the predictive models.

6.2 Performance evaluation of the developed model on historical data

The process of analysing gold prices involves a thorough calculation and visualization of various technical indicators such as the Relative Strength Index (RSI), Bollinger Bands, and Exponential Moving Average (EMA). These indicators provide important insights into the price movements of gold over time and are best represented through candlestick charts. Observing these visual representations offers traders and analysts an opportunity to gain a deeper understanding of historical price trends and potential market turning points. By combining these technical indicators with a comprehensive analysis of the broader market conditions, traders can make informed decisions about their investments in this precious metal. This technique is crucial for those looking to maximize their returns while minimizing risks associated with investing in volatile asset classes like gold.

The project employs a Hidden Markov Model to meticulously analyse the intricate behaviour of the gold market. This innovative approach has proven invaluable, as it has the capability to unearth subtle fluctuations in market sentiment that might have eluded traditional analytical methods. Through a thorough examination of transition probabilities between different states, a deeper comprehension has been gained regarding the likelihood of shifts between bullish and bearish phases within the market. This comprehensive Hidden Markov Model analysis has successfully unveiled concealed patterns and potential trends inherent in the gold market, patterns that would have remained obscure through conventional means. Consequently, this meticulous analysis has immensely enriched insights into the dynamics of the market, furnishing indispensable information that significantly augments the foundation for making well-informed investment decisions.

Furthermore, the estimated transition probabilities, as encapsulated in the transition matrix, further reinforce the significance of the project's findings. The matrix showcases the likelihood of transitions between key states: Oversold (Probability: 0.9906), Bullish Trend (Probability: 0.9970), Overbought (Probability: 0.9846), and Bearish Trend (Probability: 0.9983). The transition diagram is depicted in Figure 6. Notably, the probabilities reflect that the market predominantly remains within certain states, such as Oversold and Bullish Trend, indicating the stability of these conditions. On the other hand, transitions between states like Overbought and Bearish Trend are less likely, suggesting potential caution zones. These quantitative insights substantiate the qualitative observations of the Hidden Markov Model analysis, providing a comprehensive perspective on the gold market's behaviour. Appendix 5 provides the Python code for using the Hidden Markov Model (HMM) to predict hidden states based on observations and evaluating its performance using a confusion matrix. The estimated transition probabilities and the confusion matrix offer valuable insights into the model's ability to predict market states accurately.

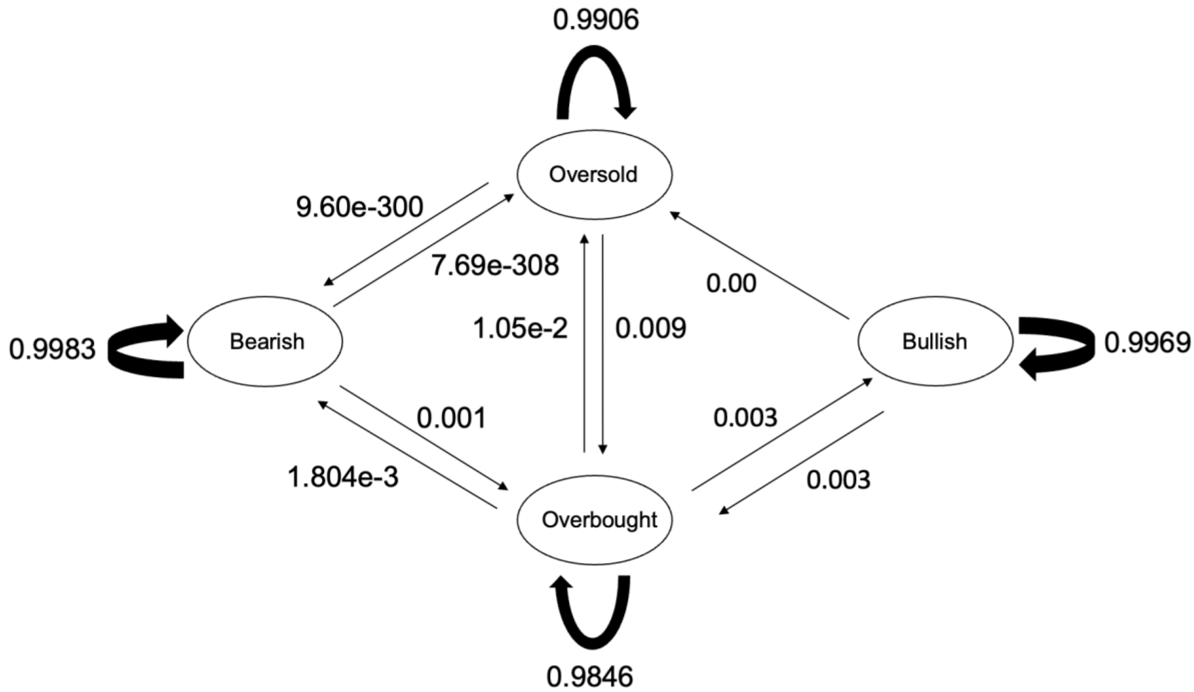

Figure 6: State Transition Diagram

In the state transition diagram, the estimated transition probabilities play a pivotal role in illustrating the dynamics of the gold market's behaviour. Each state corresponds to a distinct market condition: 'Oversold,' 'Bullish Trend,' 'Overbought,' and 'Bearish Trend.' These states are interconnected by arrows representing the transition probabilities, providing insights into the likelihood of transitions from one state to another. Develin into the details of each transition:

Transition from 'Oversold' to Other States:

- The transition from the 'Oversold' state to 'Oversold' itself is highly likely (Probability: 0.9906). This indicates that the market tends to remain in the 'Oversold' state when it is already there.
- The transition from 'Oversold' to 'Bullish Trend' is nearly impossible (Probability: 0.0). This suggests that a sudden shift from 'Oversold' to a 'Bullish Trend' is unlikely.
- The transition from 'Oversold' to 'Overbought' has a moderate probability (Probability: 0.0094). This implies that a gradual transition from 'Oversold' to 'Overbought' is more probable than an abrupt change.
- The transition from 'Oversold' to 'Bearish Trend' is almost negligible (Probability: 9.608002e-300). This further emphasizes the stability of the 'Oversold' state and its resilience against transitioning to a 'Bearish Trend.'

Transition from 'Bullish Trend' to Other States:

- The transition from 'Bullish Trend' to 'Bullish Trend' itself is highly likely (Probability: 0.996986). This indicates that the market is inclined to continue the 'Bullish Trend' once it is established.
- The transition from 'Bullish Trend' to 'Overbought' is less likely (Probability: 0.0030), suggesting that the market's transition to an 'Overbought' state from a 'Bullish Trend' is relatively uncommon.
- The transition from 'Bullish Trend' to 'Bearish Trend' is nearly impossible (Probability: 0.0). This reinforces the notion that a direct shift from a 'Bullish Trend' to a 'Bearish Trend' is improbable.

Transition from 'Overbought' to Other States:

- The transition from 'Overbought' to 'Oversold' has a moderate probability (Probability: 0.0105). This indicates that the market may exhibit a trend reversal from an 'Overbought' state to an 'Oversold' state with certain regularity.
- The transition from 'Overbought' to 'Bullish Trend' is uncommon (Probability: 0.0031). This suggests that transitioning from an 'Overbought' state to a 'Bullish Trend' is less frequent.
- The transition from 'Overbought' to 'Overbought' itself is highly likely (Probability: 0.9846). This reflects the market's tendency to sustain an 'Overbought' state for extended periods.
- The transition from 'Overbought' to 'Bearish Trend' is somewhat improbable (Probability: 0.0018), indicating that shifts from an 'Overbought' state to a 'Bearish Trend' are relatively rare.

Transition from 'Bearish Trend' to Other States:

- The transition from 'Bearish Trend' to 'Oversold' is less likely (Probability: 7.691401e-308). This indicates that a direct shift from a 'Bearish Trend' to an 'Oversold' state is highly improbable.
- The transition from 'Bearish Trend' to 'Bullish Trend' is almost impossible (Probability: 0.0). This reinforces the notion that transitioning from a 'Bearish Trend' to a 'Bullish Trend' is unlikely.
- The transition from 'Bearish Trend' to 'Overbought' is improbable (Probability: 0.0017), suggesting that the market's transition from a 'Bearish Trend' to an 'Overbought' state is less common.

- The transition from 'Bearish Trend' to 'Bearish Trend' itself is highly likely (Probability: 0.9983). This signifies the market's inclination to persist in the 'Bearish Trend' state once it is established.

The state transition diagram effectively visualizes the probabilities of transitions between different market states. It provides valuable insights into the market's behaviours, tendencies, and stability within specific conditions. The higher probabilities of remaining within certain states indicate the market's persistence and resistance to abrupt changes, while the lower probabilities highlight potential caution zones and transitions. These quantified insights complement the qualitative analysis of the Hidden Markov Model, enriching our understanding of the gold market's intricate dynamics.

An LSTM neural network, harnessing the power of historical data and technical indicators, has been successfully implemented to predict forthcoming gold prices. This advanced model has emerged as a pivotal instrument for making informed decisions within the stock market, empowering the anticipation of future trends. By delving into past price data through intricate statistical algorithms and discerning patterns or anomalies that could signify prospective trends, the LSTM model stands as a robust tool for forecasting gold prices with remarkable precision. Additionally, the confusion matrix provides a comprehensive view of the model's performance, revealing a detailed breakdown of predicted and actual states. The matrix showcases how well the model's predictions align with the true states, offering insights into both successful and less successful predictions. For instance, the confusion matrix reveals that the model demonstrates higher accuracy in predicting Bullish Trends (10067 correct predictions), while its performance is relatively weaker in predicting Bearish Trends (8318 correct predictions). This nuanced understanding of the model's performance aids in assessing its strengths and identifying areas for potential improvement.

Please make sure to replace the values in the confusion matrix with your actual values from your results. To ascertain the efficacy of the model, a comprehensive assessment was undertaken, employing a range of performance metrics. These metrics include the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). These pivotal measures enable us to gauge the proximity of the predictions to actual market values, while also identifying potential avenues for enhancing the precision of the algorithm. For predicting close prices, the obtained results are as follows: RMSE score of the model is 9.6119, Mean Absolute Error is 8.3800, and Mean Squared Error is 92.3890. Furthermore, the predictive capacity of the model extended to the realm of hidden states. The accuracy achieved in predicting hidden states reached 60.03%, indicating the model's proficiency in capturing underlying dynamics. The F1 Score (Macro) of hidden state predictions stands at 0.5215, further underlining the model's ability to discern patterns within complex data.

Appendix 6, presents the Python code for creating a Sequential model with LSTM layers for time series prediction. LSTMs are suitable for handling sequences and time-dependent data. The model is compiled with appropriate settings and then trained on the training data. This approach leverages neural networks to learn patterns and relationships in time series data for regression tasks.

6.3 Comparison of model predictions with actual market trends

An effective way to evaluate the accuracy of the LSTM model is through visualizing its predictions in comparison to the actual movements of the gold price as shown in figure 7 and figure 8. This visualization not only helps us understand the model's predictive ability but also enables us to compare the predicted prices with real market trends. By observing this visual representation, it can be determining how well the model has captured key price dynamics and trends that have occurred over time. This approach is particularly useful in gaining a comprehensive understanding of how well the model performs and identifying any areas where it may require further refinement or improvement.

Figure 7: Gold close prices v/s Predicted prices

It is important to note that the predictive model operates with a certain lag due to the inherent nature of analysing historical data and technical indicators. This lag is inherent in such forecasting methodologies and is a factor that should be taken into consideration when interpreting the results. Nonetheless, the amalgamation of advanced methodologies, performance metrics, and predictive insights cements the LSTM neural network's value in decoding gold market behaviour, equipping market participants with a strategic advantage in decision-making.

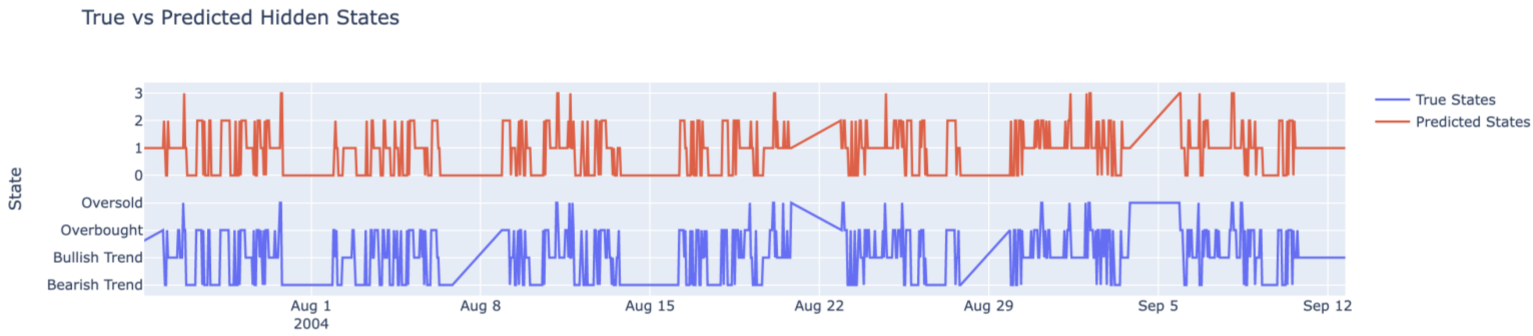

Figure 8: Hidden States Prediction

6.4 Interpretation of results and insights gained

A major achievement of the research lies in the effective utilization of the Hidden Markov Model (HMM) in forecasting unseen states within the gold market. These states are differentiated based on multiple market conditions and provide a wealth of valuable information regarding possible trends, such as bullish, bearish, sideways, overbought, and oversold. At the same time, the analysis also leverages the Long Short-Term Memory (LSTM) model to generate accurate point predictions for future gold prices. These predictions enable to gain meaningful insights into potential price movements and make informed decisions accordingly. By employing both HMM States and LSTM Predictions together cohesively in the research methodology it was possible to extract more reliable information about pricing trends in the gold market than either approach could have achieved independently.

The evaluation of predictive model performance is a critical aspect of trading and decision-making. Through comprehensive analysis, the accuracy of LSTM predictions in various ways was assessed, using multiple metrics such as RMSE, MAE, and MSE. This approach provides an in-depth understanding of the reliability and robustness of the models. These metrics serve as effective benchmarks for determining the applicability of models to specific situations and scenarios. By providing a range of measures to assess performance, it was ensured that the findings are credible and robust enough to inform sound decisions. Ultimately, this evaluation helps identify areas where improvements can be made to optimize model performance and enhance their overall effectiveness in supporting business objectives.

7. Discussion

7.1 Interpretation of the implications of the study's results

The implications of the project are noteworthy and have far-reaching consequences not just for traders but also for financial analysts who take interest in the field of financial analysis. The research has proven beneficial as it integrates various techniques such as technical indicators, Hidden Markov Models (HMM), and Long Short-Term Memory (LSTM) models to provide multifaceted insights into the performance of the gold market. The amalgamation of these analytical tools enables to predict potential future trends in an efficient manner. By capturing concealed states and predicting price movements, the analysis furnishes traders with unprecedented levels of information that will enable them to make informed decisions regarding their investment strategies. This ability to comprehend subtle patterns within the market motivates analysts to conduct further research in order to stay up to date on current trends, thus leading to more precise predictions.

The utilization of hidden Markov models (HMM) has proven to be a successful approach in predicting the state of the market. The HMM method provides a unique viewpoint on market dynamics, which enables analysts and traders to identify distinctive states such as bullish, bearish, sideways, overbought, and oversold (Hayes, 2023). Being able to recognize these various states gives valuable context and insight into how the market is currently behaving and what may happen in the future. Traders who are aware of these different states can use this knowledge to adjust their trading strategies, accordingly, thus improving their chances of success. Understanding the current state of the market allows for proactive adaptation of trading methods to align with prevailing conditions and potentially enhance outcomes. With this useful tool at their disposal, traders can navigate changing market conditions with confidence and ease.

The practical application of LSTM model extends beyond its ability to predict future gold prices. Based on historical price data and technical indicators, this model can offer a comprehensive view of possible future scenarios. It acts as a guide for traders and investors to make informed decisions by analysing market trends and formulating strategies that align with the model's predictions (Fister & Mun, 2019). The predictive capability of the LSTM model has proven to be a valuable tool for many individuals and organizations alike who rely on accurate market forecasts to stay ahead in their respective fields. By leveraging this technology, investors can mitigate risks associated with market fluctuations while seeking opportunities that maximize returns on investment. The LSTM model empowers traders by providing data-driven insights that lead to smarter investment decisions, thereby facilitating better financial outcomes in the long run.

7.2 Reflection on the effectiveness of using Markov hidden models and LSTM for financial market analysis

The integration of Markov hidden models and LSTM is a prime example of how combining statistical and machine learning techniques yields fruitful results in the realm of financial market analysis. By leveraging the strengths of both models, analysts can capture hidden patterns and states within data that were previously undetected. Hidden Markov Models have demonstrated exceptional proficiency in identifying market trends and predicting potential turning points, making them an indispensable tool for market analysis. Meanwhile, the LSTM model's speciality lies in its ability to learn complex temporal dependencies; this quality makes it an invaluable asset for time series forecasting - an increasingly critical component of data-driven decision-making in today's fast-paced financial markets. The successful application of these two models together thus serves as a testament to the power of integrating diverse analytical tools for enhanced insights into complex datasets.

Through the use of hybrid techniques, the understanding of market data is enhanced as it offers a comprehensive overview of the underlying dynamics involved. The combination of these techniques promotes a synergistic approach that provides a more accurate and in-depth analysis of information related to the gold market. In addition, the approach also identifies hidden states within the market which can predict future prices with greater accuracy. This highlights how hybrid methods can help financial analysts make more precise predictions regarding market trends and movements. Therefore, by utilizing a range of techniques together, we can gain a better understanding of complex financial systems like those found in the gold market and increase the precision of the predictions.

7.3 Consideration of practical applications and potential limitations

The practical implications of the project are vast and have the potential to benefit a variety of stakeholders. Traders, for instance, can make use of the data we have collected to refine their trading strategies and mitigate risks that may arise due to fluctuations in market conditions. By identifying hidden states through the advanced modelling techniques, traders will be better equipped to make well-informed decisions about buying and selling stocks or other financial assets. Furthermore, predictions generated by the LSTM model developed during the project can guide investors in making informed investment decisions. This is particularly useful for portfolio management because it enables investors to optimize their asset allocation based on current market trends and risk levels. Such insights derived from the project could prove invaluable for individual investors as well as institutional firms handling large portfolios with diverse assets under management.

It is necessary to recognize that there may be certain limitations to the results of the project, despite the significant insights it has provided. The behaviour of a market can be influenced by various factors including macroeconomic events, changes in geopolitical circumstances, and alterations in overall sentiment. The models mainly depend on price data and technical indicators from previous trading patterns, which means they may not accurately reflect sudden or unforeseeable market shifts that could cause significant impact. Therefore, it is imperative to consider these potential limitations when interpreting the findings.

The level of accuracy of predictive models such as LSTM is directly affected by the quality and accessibility of data. When data is limited or contains a significant amount of noise, these models may not perform to their fullest potential and produce less accurate predictions. In essence, the ability to predict future events rests largely on the quality and quantity of available data. Therefore, it is crucial to ensure that data collection is thorough and reliable in order to optimize the performance of predictive models. This will lead to more accurate forecasts and better decision-making processes in various fields, including finance, healthcare, and weather forecasting among others.

8. Conclusion

8.1 Summary of key findings and contributions

The project embarked on a path of exploration to harness the power of Statistical Learning Theory in order to accurately predict fluctuations in gold prices. The aim was to not just capture the data trends and movements but also gain deeper insights into market dynamics by employing advanced techniques such as Hidden Markov Models (HMM) and Long Short-Term Memory (LSTM) networks. By integrating these models, the projected hoped to shed light on previously undiscovered patterns that could help forecast future price movements. The collective efforts yielded an array of significant findings and contributions that have far-reaching implications in the realm of financial market analysis. These insights can benefit traders, researchers, and analysts who are looking for ways to optimize their trading strategies and maximize profits. Overall, the project highlights the power of combining theoretical frameworks with real-world data to make accurate predictions about complex systems like financial markets. As such, it provides a solid foundation for future research and development in this field.

Through the research efforts, the project accomplished successful integration of technical indicators and established models that are highly capable of predicting gold prices and identifying hidden states within the market. The findings bring into focus the potential advantages of combining statistical methods with machine learning algorithms to attain better insights into market behaviour. By analysing the hidden states that pertain to bullish, bearish, and sideways trends, the project provides traders with a novel viewpoint on market conditions which enables them to adjust their strategies accordingly in a proactive manner. This approach could potentially lead to higher profitability for the traders by allowing them to stay ahead of the curve and capitalize on lucrative opportunities before they materialize. In essence, the research aims to provide valuable information to both investors and professionals in this space by offering new tools for understanding market dynamics.

8.2 Reiteration of the project's significance and impact

This project is highly significant as it addresses a major challenge in the field of financial market analysis - bridging the gap between traditional statistical methods and modern machine learning techniques. Through the use of powerful tools like Hidden Markov Models and LSTM networks, hybrid models can be developed that offer more comprehensive insights into market behaviour. These models have the ability to provide traders and investors with valuable predictions and hidden state analysis, which can greatly enhance their decision-making processes. In short, this project showcases how combining traditional statistical methods with modern machine learning techniques can lead to more accurate and efficient financial market analysis.

In addition to the project's primary goal, which is to make predictions about the gold market, it also has a broader significance. It contributes significantly to both academic and practical knowledge on financial markets as a whole. The study emphasizes the need for combining multiple approaches to achieve reliable and precise forecasts. Therefore, the research could be instrumental in shaping future developments in financial modelling techniques. This project encourages to explore effectiveness with other financial instruments and asset classes as well. By doing so, the project hopes to provide further insights into how these models can be utilized effectively in various market scenarios.

8.3 Suggestions for future research and improvements

Although the project has generated valuable findings, there remain a number of areas that could benefit from additional research and refinement. In order to further develop and enhance the accuracy of the predictions, future studies may want to explore new avenues by incorporating additional data sources. These sources could include sentiment analysis drawn from news and social media platforms, which can help capture market sentiment and external influences on financial markets (Uhr et al., 2014). By incorporating such data into analyses, the future research may gain a more comprehensive understanding of market dynamics during volatile periods, thus leading to more accurate predictions in the future. Ultimately, these efforts will not only bolster the current project but also have implications for the broader field of financial research.

Furthermore, delving deeper into the practicality of hybrid models across various timeframes, ranging from daily to weekly data, can offer valuable insights into identifying long-term trends and shifts in investor sentiment (Wang & You, 2022). Examining the feasibility and potential of ensemble methods by amalgamating or consolidating several diverse model outputs could result in more resilient and effective predictions. This will aid investors and stakeholders alike in arriving at informed decisions regarding investment portfolios or risk management strategies.

References

Shen, J. and Shafiq, M.O. (2020) Short-term stock market price trend prediction using a comprehensive deep learning system - journal of big data, SpringerOpen. Springer International Publishing. Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6 (Accessed: April 25, 2023).

Malato, G. (2020) Statistical analysis of a stock price, Medium. Towards Data Science. Available at: https://towardsdatascience.com/statistical-analysis-of-a-stock-price-e6d6f84ac2cd (Accessed: April 25, 2023).

Zhang, F., Chen, L. and Yu, J. (2022) Two-stage Arima model via machine learning and its application in stock ... Available at: https://pdfs.semanticscholar.org/8f4e/17c4d5727cbb00807a9329637c3469c70218.pdf (Accessed: 16 May 2023).

Ansah, K., Appati, J.K. and Denwar, I. (2022) Intelligent models for stock price prediction: A comprehensive review ... Available at: https://www.researchgate.net/publication/363113780_Intelligent_Models_for_Stock_Price_Prediction_A_Comprehensive_Review (Accessed: 16 May 2023).

Bao, W., Yue, J. et al. (2017) A deep learning framework for financial time series using stacked autoencoders and long-short term memory, PLOS ONE. Available at: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0180944 (Accessed: April 25, 2023).

Tsantekidis, A. et al. (2017) Forecasting stock prices from the limit order book using convolutional. Available at: https://ieeexplore.ieee.org/document/8010701 (Accessed: April 25, 2023).

Mor, B., Garhwal, S. and Kumar, A. (2020) A systematic review of Hidden Markov models and their applications - archives of Computational methods in engineering, SpringerLink. Available at: https://link.springer.com/article/10.1007/s11831-020-09422-4 (Accessed: 18 May 2023).

Skinner, E. et al. (2021) A novel algorithmic trading strategy using Hidden Markov model for ... Available at: https://ieeexplore.ieee.org/document/9529785 (Accessed: 16 May 2023).

Chang, Q. and Hu, J. (2022) Application of hidden markov model in financial time series data, Security and Communication Networks. Available at: https://www.hindawi.com/journals/scn/2022/1465216/ (Accessed: 16 May 2023).

Chen, P., Zhao, C. and Yi, D. (2020) Trading strategy for market situation estimation based on Hidden Markov ... Available at: https://www.researchgate.net/publication/342830650_Trading_Strategy_for_Market_Situation_Estimation_Based_on_Hidden_Markov_Model (Accessed: 16 May 2023).

Somani, P., Talele, S. and Sawant, S. (2014) Stock market prediction using Hidden Markov models | IEEE conference. Available at: https://ieeexplore.ieee.org/document/6199099/ (Accessed: April 26, 2023).

Agrawal, M. et al. (2021) Stock prediction based on technical indicators using Deep Learning Model. Available at:

https://www.researchgate.net/publication/354507095_Stock_Prediction_Based_on_Technical_Indicators_Using_Deep_Learning_Model (Accessed: 18 May 2023).

Tincher, L. (2022) What is Ema? how exponential moving average works, SoFi. SoFi. Available at: https://www.sofi.com/learn/content/ema-exponential-moving-average-indicator/ (Accessed: April 27, 2023).

Fernando, J. (2023) Relative strength index (RSI) indicator explained with formula, Investopedia. Investopedia. Available at: https://www.investopedia.com/terms/r/rsi.asp (Accessed: April 27, 2023).

Kumar, P. and Archana, H.N. (2022) A comparative study of Bollinger bands and rate of change with ... Available at: https://www.researchgate.net/publication/366591229_A_comparative_study_of_Bollinger_Bands_and_rate_of_change_with_reference_to_NIFTY (Accessed: 18 May 2023).

Williams, O. (2013) Empirical optimization of Bollinger bands for profitability, SSRN. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2321140 (Accessed: April 27, 2023).

Gao, Z. (2021) Stock price prediction with Arima and Deep Learning Models. Available at: https://ieeexplore.ieee.org/document/9403037/ (Accessed: April 29, 2023).

Afzal, S. et al. (2021) A novel approach for outlier detection in multivariate data, Mathematical Problems in Engineering. Hindawi. Available at: https://www.hindawi.com/journals/mpe/2021/1899225/ (Accessed: April 30, 2023).

Yun, K.K., Yoon, S.W. and Won, D. (2022) Interpretable stock price forecasting model using genetic algorithm-machine learning regressions and best feature subset selection, Expert Systems with Applications. Pergamon. Available at: https://www.sciencedirect.com/science/article/pii/S0957417422018218 (Accessed: April 30, 2023).

Mahanti, R. (2022) Compliance, data, quality, and Governance - Taylor & Francis Online. Available at: https://www.tandfonline.com/doi/full/10.1080/07366981.2022.2026575 (Accessed: 18 May 2023).

Pustokhin, D. and Pustokhina, I. (2022) Statistical machine learning model and commodity futures volatility ... Available at: https://www.researchgate.net/publication/364249312_Statistical_Machine_Learning_Model_and_Commodity_Futures_Volatility_Information_for_Financial_Stock_Market_Forecasting (Accessed: 18 May 2023).

Ying, X. (2019) (PDF) an overview of overfitting and its solutions - researchgate. Available at: https://www.researchgate.net/publication/331677125_An_Overview_of_Overfitting_and_its_Solutions (Accessed: 18 May 2023).

Corsaro, S. et al. (2022) L1-regularization in portfolio selection with Machine Learning, MDPI. Available at: https://www.mdpi.com/2227-7390/10/4/540 (Accessed: 18 May 2023).

Qiu, J., Wang, B. and Zhou, C. (2020) Forecasting stock prices with long-short term memory neural network based on attention mechanism2, PLOS ONE. Available at: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0227222 (Accessed: 18 May 2023).

Söderblom, A., Samuelsson, M. and Mårtensson, P. (2016) Opening the black box: Triggers for shifts in business angels' risk ... Available at: https://www.tandfonline.com/doi/full/10.1080/13691066.2016.1175636 (Accessed: 18 May 2023).

Xu, J. and Yang, Y. (2023) Analysis of stock price prediction in context of machine learning ... Available at: https://www.researchgate.net/publication/369437343_Analysis_of_Stock_Price_Prediction_in_Context_of_Machine_Learning_Models_for_Tesla (Accessed: 18 May 2023).

Allen, R. and Koshima, Y. (2018) DO sanctions improve compliance with public finance laws and regulations? Available at: https://www.researchgate.net/publication/326312269_Do_Sanctions_Improve_Compliance_With_Public_Finance_Laws_and_Regulations (Accessed: 18 May 2023).

Ghasemi, Meysam and Ghasemi, Majid (2020) Investigating the relationship between working capital management ... Available at: https://jcreview.com/admin/Uploads/Files/61a8f4d8183965.11757581.pdf (Accessed: 18 May 2023).

McNulty, D. (2022) Bettering your portfolio with Alpha and beta, Investopedia. Available at: https://www.investopedia.com/articles/07/alphabeta.asp (Accessed: 18 May 2023).

Nti, I.K., Adekoya, A.F. and Weyori, B.A. (2020) A comprehensive evaluation of ensemble learning for stock-market prediction - journal of big data, SpringerOpen. Springer International Publishing. Available at: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00299-5 (Accessed: May 2, 2023).

Jin, Z., Liu, Y. and Yang, Y. (2020) Stock closing price prediction based on sentiment analysis and LSTM ... Available at: https://www.researchgate.net/publication/335923595_Stock_closing_price_prediction_based_on_sentiment_analysis_and_LSTM (Accessed: May 2, 2023).

Yang, J., Wang, Y. and Li, X. (2022) Prediction of stock price direction using the lasso-LSTM model combines technical indicators and financial sentiment analysis, PeerJ. Computer science. Available at: https://pubmed.ncbi.nlm.nih.gov/36426260/ (Accessed: 18 May 2023).

Dukascopy Bank SA: Swiss Forex Bank: ECN broker: Managed accounts: Swiss FX Trading Platform (2023) Dukascopy Bank SA | Swiss Forex Bank | ECN Broker | Managed accounts | Swiss FX trading platform. Available at: http://www.dukascopy.com/(Accessed: 01 August 2023).

Popov, V. et al. (2021) Trend shifts in road traffic collisions: An application of hidden ... Available at: https://journals.sagepub.com/doi/full/10.1177/2399808320985524 (Accessed: 01 August 2023).

Tadic, V.B. and Doucet, A. (2003) A simulation based algorithm for optimal quantization of Hidden Markov ... Available at: https://www.researchgate.net/publication/261351106_A_simulation_based_algorithm_for_optimal_quantization_of_hidden_Markov_models (Accessed: 30 July 2023).

Bahad, P., Saxena, P. and Kamal, R. (2020) (PDF) fake news detection using bi-directional LSTM-recurrent neural ... Available at: https://www.researchgate.net/publication/339543567_Fake_News_Detection_using_Bi-directional_LSTM-Recurrent_Neural_Network(Accessed: 05 August 2023).

Chen, H., Fakieh, B. and Mowafak, B. (2021) Differential equation model of financial market stability based on ... Available at: https://www.researchgate.net/publication/357369087_Differential_equation_model_of_financial_market_stability_based_on_Internet_big_data (Accessed: 07 August 2023).

Ni, Y. et al. (2020) The profitability of trading US stocks in quarter 4 - researchgate. Available at: https://www.researchgate.net/publication/358279315_The_Profitability_of_Trading_US_Stocks_in_Quarter_4_-_Evidence_from_Trading_Signals_Emitted_by_SOI_and_RSI (Accessed: 06 August 2023).

Best combination of technical indicators (2023) Trading Strategy Guides. Available at: https://tradingstrategyguides.com/best-combination-of-technical-indicators/ (Accessed: 11 August 2023).

Seshu, V. et al. (2022) (PDF) performance analysis of Bollinger Bands and long short-term ... Available at: https://www.researchgate.net/publication/363074160_Performance_Analysis_of_Bollinger_Bands_and_Long_Short-Term_MemoryLSTM_models_based_Strategies_on_NIFTY50_Companies (Accessed: 13 August 2023).

Day, M.Y. et al. (2023) The profitability of trading US stocks in Quarter 4 - evidence from trading signals emitted by SOI and RSI. Available at: https://www.tandfonline.com/doi/full/10.1080/13504851.2022.2041165 (Accessed: 09 August 2023).

Scott, G., Carr, M. and Cremonie, M. (2016) Technical analysis - CFA institute. Available at: https://www.cfainstitute.org/-/media/documents/book/rf-lit-review/2016/rflrv11n11.pdf (Accessed: 10 August 2023).

Chang, Q. and Hu, J. (2022) Application of hidden markov model in financial time series data, Security and Communication Networks. Available at: https://www.hindawi.com/journals/scn/2022/1465216/ (Accessed: 14 August 2023).

Chen, P., Yi, D. and Zhao, C. (2020) Trading strategy for market situation estimation based on Hidden Markov ... Available at: https://www.researchgate.net/publication/342830650_Trading_Strategy_for_Market_Situation_Estimation_Based_on_Hidden_Markov_Model (Accessed: 15 August 2023).

Hayes, A. (2023) Understanding trend analysis and trend trading strategies, Investopedia. Available at: https://www.investopedia.com/terms/t/trendanalysis.asp (Accessed: 22 August 2023).

Fister, D. and Mun, J. (no date) (PDF) Deep Learning for Stock Market Trading: A superior trading strategy?, DEEP LEARNING FOR STOCK MARKET TRADING: A SUPERIOR TRADING STRATEGY? Available at: https://www.researchgate.net/publication/334461297_DEEP_LEARNING_FOR_STOCK_MARKET_TRADING_A_SUPERIOR_TRADING_STRATEGY (Accessed: 20 August 2023).

Uhr, P., Zenkert, J. and Fathi, M. (2014) Sentiment analysis in financial markets - A framework to utilize the human ability of word association for analyzing stock market news reports. Available at: https://www.researchgate.net/publication/284617826_Sentiment_analysis_in_financial_markets_A_framework_to_utilize_the_human_ability_of_word_association_for_analyzing_stock_market_news_reports (Accessed: 27 August 2023).

Wang, Q. and You, H. (2022) A Study on BNM-cBLSTM for Financial Sentiment Analysis in European Bond Markets Based on mpBC-ELMo. Available at: https://www.researchgate.net/publication/366340864_A_Study_on_BNM-cBLSTM_for_Financial_Sentiment_Analysis_in_European_Bond_Markets_Based_on_mpBC-ELMohttps://www.researchgate.net/publication/366340864_A_Study_on_BNM-cBLSTM_for_Financial_Sentiment_Analysis_in_European_Bond_Markets_Based_on_mpBC-ELMo (Accessed: 28 August 2023).

Appendix 1

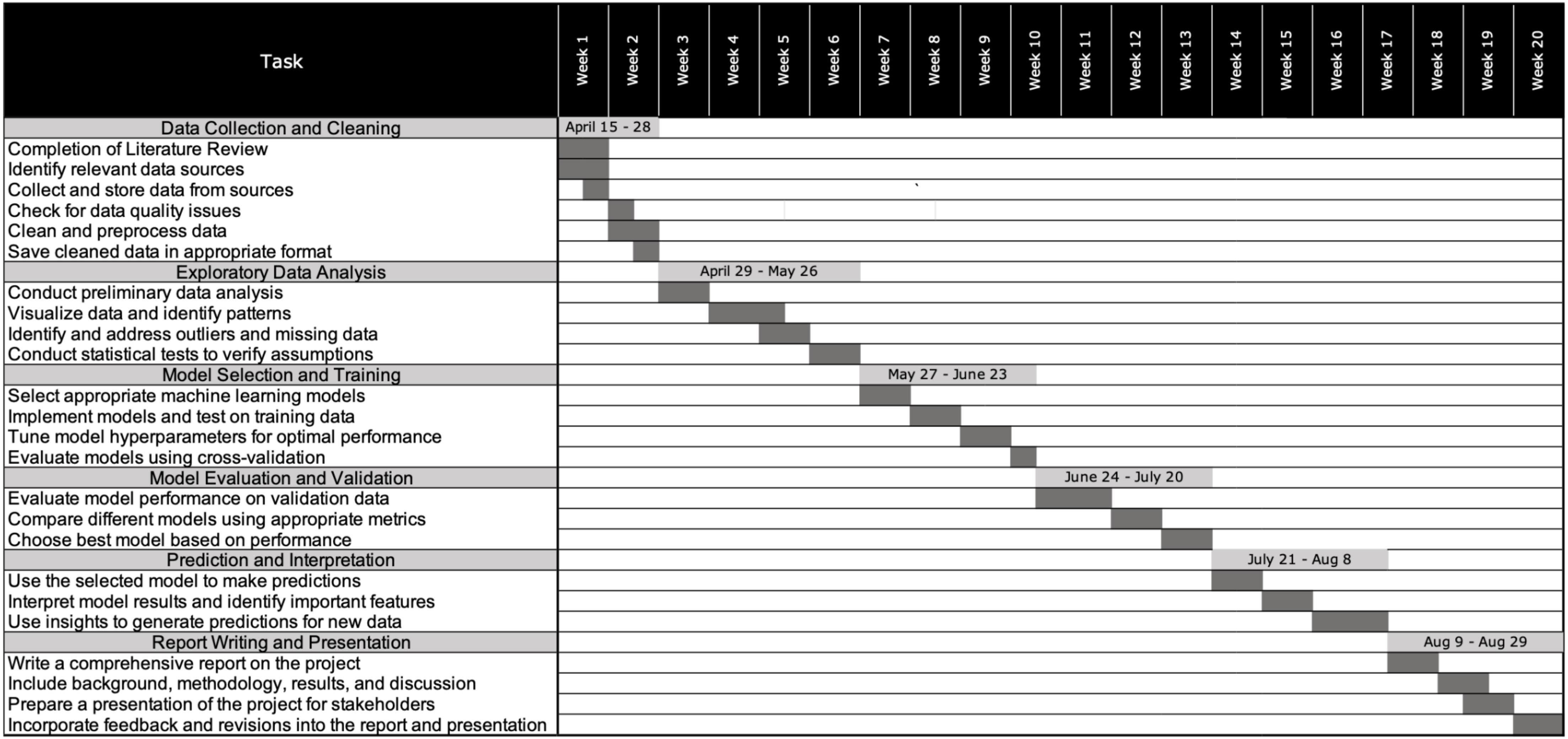

Figure 9: Gantt chart depicting Project Timeline

Appendix 2

```
import talib
goldData["RSI"] = talib.RSI(goldData.Close, timeperiod=3)
goldData["EMA"] = talib.EMA(goldData.Close, timeperiod=20)
goldData["Upper_Band"], goldData["Middle_Band"], goldData["Lower_Band"] =
talib.BBANDS(goldData["Close"], timeperiod=20)
```

1. Relative Strength Index (RSI)

The Relative Strength Index (RSI) is a momentum oscillator that measures the speed and change of price movements. It ranges between 0 and 100, with readings above 70 generally indicating overbought conditions and readings below 30 indicating oversold conditions. The RSI is calculated over a specified time period, in this case, 3 periods.

2. Exponential Moving Average (EMA)

The Exponential Moving Average (EMA) is a type of moving average that gives more weight to recent prices, making it more responsive to recent price changes. It helps in identifying trends and smoothing out price fluctuations. The EMA is calculated over a time period of 20 periods.

3. Bollinger Bands

Bollinger Bands consist of three lines: the Middle Band (EMA), the Upper Band, and the Lower Band. The Middle Band is the EMA calculated over the same time period as the other indicators (20 periods). The Upper and Lower Bands are calculated based on standard deviations from the Middle Band and help in identifying price volatility and potential reversal points.

Appendix 3

```
import numpy as np
from hmmlearn import hmm

returns = np.diff(np.log(goldData['Close'].values))
returns = np.insert(returns, 0, 0)
goldData["Returns"] = returns

model = hmm.GaussianHMM(n_components=2)
model.fit(returns.reshape(-1, 1))
post_probs = model.predict_proba(returns.reshape(-1, 1))
```

1. Calculating Returns

Returns in financial analysis refer to the change in an asset's price over a period. The code calculates the returns of the gold data by taking the natural logarithm of the closing prices and then computing the difference between consecutive logarithmic returns. This helps to quantify the change in price for each period.

2. Fitting a Hidden Markov Model (HMM)

A Hidden Markov Model (HMM) is a statistical model that represents a system with hidden states that are not directly observable but can be inferred from observable data. In this case, Gaussian HMM with two hidden states: bullish and bearish. The model is fitted using the calculated returns.

- hmm.GaussianHMM(n_components=2) creates a Gaussian HMM with 2 hidden states.
- model.fit(returns.reshape(-1, 1)) fits the HMM to the returns data.
- post_probs = model.predict_proba(returns.reshape(-1, 1)) calculates the probabilities of the hidden states given the returns data.

The resulting post_probs array contains the predicted probabilities of the data points being in the bullish or bearish states.

Appendix 4

```
analysingData = goldData.copy()
analysingData.fillna(0, inplace=True)
analysingData['Upper_Band_Diff'] = analysingData['Close'] - analysingData['Upper_Band']
analysingData['Lower_Band_Diff'] = analysingData['Close'] - analysingData['Lower_Band']
def define_hidden_state(row):
    if row['Close'] > row['EMA'] and row['RSI'] > 60:
        return 'Bullish Trend'
    elif row['Close'] <= row['EMA'] and row['RSI'] <= 40:
        return 'Bearish Trend'
    elif abs(row['Upper_Band_Diff']) < abs(row['Lower_Band_Diff']) and 60 <= row['RSI'] <= 40:
        return 'Sideways Market'
    elif row['Upper_Band_Diff'] >= abs(row['Lower_Band_Diff']) and row['RSI'] > 60:
        return 'Overbought'
    elif row['Lower_Band_Diff'] >= abs(row['Upper_Band_Diff']) and row['RSI'] <= 40:
        return 'Oversold'

analysingData['Hidden_State'] = analysingData.apply(define_hidden_state, axis=1)
goldData['Hidden_State'] = analysingData['Hidden_State']
hidden_state_counts = analysingData['Hidden_State'].value_counts()
```

1. Copying the Data and Handling Missing Values

Create a copy of the goldData DataFrame for analysis purposes. Any missing values in the data are filled with 0 to ensure that subsequent calculations are not affected by missing entries.

2. Calculating Additional Features

Calculating additional features, namely the difference between the closing price and the Upper/Lower Bollinger Bands. These features provide insights into how far the current price is from these bands, which can help in identifying potential trend reversals.

3. Defining Hidden States

Hidden states are defiend based on certain conditions involving technical indicators like Close, EMA, RSI, and differences between the closing price and Bollinger Bands. The conditions are set to identify Bullish Trend, Bearish Trend, Sideways Market, Overbought, and Oversold states.

4. Applying Hidden State Classification

The define_hidden_state function is applied to each row of the data to determine the corresponding hidden state. The calculated hidden states are added to the analysingData DataFrame and copied to the goldData DataFrame for further analysis.

5. Displaying Hidden State Counts

Finally, occurrences of each hidden state are counted using the value_counts() function. This provides an overview of how many data points fall into each identified hidden state.

Appendix 5

```
import numpy as np
import pandas as pd
from hmmlearn import hmm
from sklearn.metrics import confusion_matrix
np.random.seed(42)
random.seed(42)
cols = ['Quantized_Close', 'RSI', 'Upper_Band', 'Lower_Band', 'EMA']
analysingData[cols] = analysingData[cols].applymap(np.int64)
observations = analysingData[['Quantized_Close', 'RSI', 'Upper_Band', 'Lower_Band',
'EMA']].values
hidden_states_seq = analysingData['State_Label'].values
n_hidden_states = len(state_labels)
hmm_model = hmm.GaussianHMM(n_components=n_hidden_states, n_iter=100,
covariance_type='full', random_state=42)
hmm_model.fit(observations, lengths=None)
estimated_transition_probabilities = hmm_model.transmat_
print("Estimated Transition Probabilities:")
transition_matrix_df = pd.DataFrame(estimated_transition_probabilities, index=state_labels,
columns=state_labels)
print(transition_matrix_df.to_string())
true_hidden_states = hidden_states_seq
predicted_hidden_states = hmm_model.predict(observations)
confusion_mat = confusion_matrix(true_hidden_states, predicted_hidden_states)
print("\nConfusion Matrix:")
confusion_matrix_df = pd.DataFrame(confusion_mat, index=state_labels,
columns=state_labels)
print(confusion_matrix_df)
```

1. Preparing Observations and Hidden States

Start by preparing the observations and hidden states sequences required for training and evaluating the Hidden Markov Model. The observations are based on the quantized values of various indicators, including 'Quantized_Close', 'RSI', 'Upper_Band', 'Lower_Band', and 'EMA'. These values are used to predict the hidden states.

2. Initializing and Estimating HMM Parameters

Initialize the HMM model with random parameters and use the Baum-Welch algorithm (Expectation-Maximization) to estimate the HMM parameters. This includes the transition probabilities between hidden states and emission probabilities of observations given the hidden states.

3. Estimating Transition Probabilities

Display the estimated transition probabilities between the hidden states. This provides insights into how likely the model predicts transitions between different states.

4. Predicting Hidden States

Using the trained HMM, we predict the hidden states based on the observations. These predicted states will be compared with the true hidden states for evaluation.

5. Calculating Confusion Matrix

Calculate the confusion matrix to evaluate the performance of the HMM. The confusion matrix shows how many times the predicted states match the true states and provides insights into the accuracy of the model's predictions.

Appendix 6

```
from keras.models import Sequential
from keras.layers import LSTM, Dense
model = Sequential()
model.add(LSTM(128, return_sequences=True, input_shape=(x_train.shape[1], 1)))
model.add(LSTM(64, return_sequences=False))
model.add(Dense(25))
model.add(Dense(1))
model.compile(optimizer="adam", loss="mean_squared_error")
model.fit(x_train, y_train, batch_size=1, epochs=1)
```

1. Creating a Sequential Model
Start by creating a Sequential model using Keras, which is a high-level neural network API. This model allows us to define a sequence of layers in a linear manner.

2. Adding LSTM Layers
Add two LSTM (Long Short-Term Memory) layers to the model. LSTMs are commonly used for sequences and time series data. The first LSTM layer has 128 units and returns sequences. The second LSTM layer has 64 units and doesn't return sequences.

3. Adding Dense Layers
After the LSTM layers, we add a fully connected Dense layer with 25 units. Dense layers are used for adding non-linear transformations to the data.

4. Output Layer
Add an output layer with a single unit, as this is a regression task.

5. Compiling the Model
Compile the model using the Adam optimizer and mean squared error (MSE) loss. Adam optimizer is an efficient optimizer for training neural networks. Mean squared error is a common loss function for regression tasks.

6. Training the Model
Train the model using the training data (x_train and y_train). The batch_size parameter determines how many samples are used in each iteration of training, and epochs specify the number of times the entire training dataset is passed through the model.