PREDICTION OF GOLD PRICES BASED ON M5P MODEL TREE AND LSTM WITH MULTI FACTORS ANALYSIS

LOH CIN CEAT

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2023

PREDICTION OF GOLD PRICES BASED ON M5P MODEL TREE AND LSTM WITH MULTI FACTORS ANALYSIS

LOH CIN CEAT

RESEARCH REPORT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE

FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
KUALA LUMPUR

2023

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **LOH CIN CEAT** (IC/Passport No: **980706-05-5107**)

Matric No: **S2141070**

Name of Degree: **MASTER OF DATA SCIENCE**

Title of Project Paper ("this Work"):

PREDICTION OF GOLD PRICES BASED ON M5P MODEL TREE AND LSTM WITH MULTI FACTORS ANALYSIS

Field of Study:

MACHINE LEARNING AND DEEP LEARNING

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: 03-Feb-2023

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

PREDICTION OF GOLD PRICES BASED ON M5P MODEL TREE AND LSTM WITH MULTI FACTORS ANALYSIS

ABSTRACT

For centuries, gold has been a highly valued commodity and is regarded as a secure asset. It is used for luxury and jewelry purposes, as well as in electronics and high-tech industries. This article explores the influence of economic variables on gold prices, such as macroeconomic indicators, stock indexes and other commodities. Machine learning models (M5P model trees) and deep learning models (LSTM) were employed to predict gold prices 1-day and 7-days ahead. The economic variables are studied for their correlation with gold price and are used as input to predict the gold price. Furthermore, technical indicators such as Bollinger Band, SMA, MACD, RSI and ROC were used to evaluate their efficacy in forecasting gold prices. The results revealed that US inflation rates, silver prices, copper prices, NASDAQ, DJI, and S&P500 indexes are strongly positively correlated with the gold price. The least positively correlated variable was Brent Crude Oil prices. On the flip side, three macroeconomic factors: the US Dollar Index, interest rates, and US Treasury Bond Yield have inverse relationship with the gold price. The prediction results shows that technical indicators had higher predictive power than economic indicators; moreover, their accuracy was not compromised by issues related to data mining or time changes. Our findings suggest that both the M5P and LSTM models can accurately predict gold prices 1-day and 7 days ahead with R^2 values above 0.97 when using technical analysis as inputs.

Keywords: Gold prices, Machine learning, Deep learning, Influence factors, Technical indicators.

PREDICTION OF GOLD PRICES BASED ON M5P MODEL TREE AND LSTM WITH MULTI FACTORS ANALYSIS

ABSTRAK

Selama berabad-abad, emas telah menjadi komoditi yang sangat bernilai dan dianggap sebagai aset yang selamat. Ia digunakan untuk tujuan mewah dan barang kemas, serta dalam industri elektronik dan teknologi tinggi. Artikel ini meneroka pengaruh pembolehubah ekonomi ke atas harga emas, seperti penunjuk makroekonomi, indeks saham dan komoditi lain. Model pembelajaran mesin (pokok model M5P) dan model pembelajaran mendalam (LSTM) digunakan untuk meramalkan harga emas 1 hari dan 7 hari ke hadapan. Pembolehubah ekonomi dikaji untuk korelasinya dengan harga emas dan digunakan sebagai input untuk meramalkan harga emas. Tambahan pula, penunjuk teknikal seperti Bollinger Band, SMA, MACD, RSI dan ROC digunakan untuk menilai keberkesanannya dalam meramalkan harga emas. Keputusan menunjukkan bahawa kadar inflasi AS, harga perak, harga tembaga, indeks NASDAQ, DJI, dan S&P500 sangat berkorelasi positif dengan harga emas. Pembolehubah yang paling kurang berkorelasi positif ialah harga Minyak Mentah Brent. Sebaliknya, tiga faktor makroekonomi: Indeks Dolar AS, kadar faedah, dan Hasil Bon Perbendaharaan AS mempunyai hubungan songsang dengan harga emas. Keputusan ramalan menunjukkan penunjuk teknikal mempunyai kuasa ramalan yang lebih tinggi daripada penunjuk ekonomi; selain itu, ketepatannya tidak terjejas oleh isu yang berkaitan dengan perlombongan data atau perubahan masa. Penemuan kami mencadangkan bahawa kedua-dua model M5P dan LSTM boleh meramalkan harga emas dengan tepat 1 hari dan 7 hari ke hadapan dengan nilai R^2 melebihi 0.97 apabila menggunakan analisis teknikal sebagai input.

Keywords: Harga emas, Pembelajaran mesin, Pembelajaran mendalam, Faktor pengaruh, Penunjuk teknikal.

ACKNOWLEDGEMENTS

Words cannot express my gratitude to my supervisor for her invaluable patience, feedback and guidance. I am also grateful to my classmates for their assistance and moral support. Lastly, I would be remiss in not mentioning my family, especially my parents. Their belief in me has kept my spirits and motivation high during this process.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

ADAM	:	Adaptive Moment Estimation
AIC	:	Akaike's Information Criterion
ANN	:	Artificial Neural Networks
ARIMA	:	Autoregressive Integrated Moving Average
CNN	:	Convolutional Neural Networks
DJIA	:	Dow Jones Industrial Average
DOLS	:	Dynamic Least Squares
EMA	:	Exponential Moving Average
GARCH	:	Moving Average Convergence Divergence
GDP	:	Central canal
LSTM	:	Long Short-term Memory
MA	:	Moving Average
MACD	:	Moving Average Convergence Divergence
MAE	:	Mean Absolute Error
MAPE	:	Mean Absolute Percentage Error
MLP	:	Multi-Layer Perceptron
MSE	:	Mean Absolute Error
R^2	:	Coefficient of determination
RMSE	:	Root Mean Square Error
ROC	:	Rate of Change
RSI	:	Relative Strength Index
S&P500	:	Standard & Poor's 500 Index
SMA	:	Simple Moving Average
SMA	:	Generalised Autoregressive Conditional Heteroskedasticity

SVM	:	Support Vector Machines
US	:	United States
USDX	:	United States Dollar Index

CHAPTER 1: INTRODUCTION

1.1 Research Background

Gold is one of the most significant commodities known to mankind for ages, traded on global financial markets. It is viewed as an asset that, unlike currencies, preserves its actual value. Investors have flocked to gold assets during market downturns because gold is seen as a secure asset compared to currencies due to its physical value. Apart from its financial value as an investment asset, gold is widely employed in luxury and jewelry sectors, as well as in the electronics and high-tech industries, where it is used as a high-grade conducting material in the fabrication of printed circuit boards and semiconductor components (Patalay and Bandlamudi, 2021).

The fluctuation of gold prices depends on changes in several influencing factors, where according to Liya *et al.* (2021), there is a significant link between macroeconomic indicators such as real-world GDP, inflation rates, and standard trade value with the gold. Furthermore, there is a negative correlation between the gold price and the US treasury bond yield, where treasury bond yield is the percentage return on investment on the US government's debt obligations, and gold and treasury bonds are regarded as safe-haven investments. Hence, there is a negative correlation between gold and bond rates (SunshineProfits.com, 2022).

Moreover, major US stock indexes, including S&P500, Dow Jones Industrial Average (DJIA) and Nasdaq Composite may also impact the gold price. According to Liu & Li (2017), gold investment may effectively balance the loss of stock investment due to the function of risk hedging. Hence, there is a strong negative association between the price of gold and stock indexes.

Furthermore, the other commodities market influences gold price as a commodity product. The investigation of Liu et al. (2017) shows a high correlation between copper prices and crude oil, gold and silver prices. Furthermore, it is found that the price of copper is highly correlated with the gold price since gold and copper are mined in comparable methods and have equivalent production cost structures. The copper/gold ratio is used as a "growth-to-fear proxy" when the global economy is robust, copper will rise due to the rising demand for copper in industrial activity. In contrast, the gold price will increase when there is a high fear level, such as during a recession (Bishop, 2020).

Although the variables above correlate with the gold price change, only several factors are focused on in this research. The study interest for macroeconomics variables in this research is the US dollar index (USDX), interest rates (Federal Funds Effective Rate/ Fed Fund Rates), inflation rates (Consumer Price Index), and US treasury bond yield (United States 10-Year Bond Yield). Moreover, from the ranges of major US stock indexes, Standard & Poor's 500 Index (S&P500), NASDAQ Composite Price and Dow Jones Industrial Average Price were chosen. Furthermore, commodities prices include silver, copper and oil prices. In summary, there are 10 factors affecting the gold price studied in this research: the US dollar index, interest rates, inflation rates, US treasury bond yield, SP500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil price.

Besides the factors affecting gold price mentioned above, this research has incorporated technical indicators for predicting gold price direction. Technical indicators were primarily used by technical analysts who would use the indicators to predict trade's future movement to assist in making a buy/sell decision. Much evidence from (Wang, Liu and Wu, 2020; Sadorsky, 2021;Yin and Yang, 2016) proves that technical analysis is suitable for predicting commodity prices. Several well-known technical indicators such as Bollinger Band, SMA (Simple Moving Average), MACD (Moving Average

Convergence Divergence), RSI (Relative Strength Index), and ROC (Rate of Change) are included.

In this research, tree-based machine learning models, the M5P model tree and one deep learning model, the long short-term memory (LSTM) model, were built and compared for the model's accuracy. The M5P model tree is a decision tree-based model with linear regression functions in the leaf nodes (Blaifi *et al.*, 2018). The decision tree represents the algorithms in the form of a tree that has been trained using data to produce nodes. The decision tree's nodes are classified into leaf nodes, root nodes, and internal nodes. Nodes are linked to one another from branches until reaching the leaves (Wang and Witten, 1997).

On the other hand, LSTM is a type of artificial neural network that handles time-series or sequential data (Khan, Ranjan and Kumar, 2021). LSTM is a type of recurrent neural network which more competent in finding patterns from huge dataset sequences. LSTM is also known to respond better to non-linearity (Mohtasham Khani, Vahidnia and Abbasi, 2021).

1.2 Problem Statement

Multiple linear regression models, panel data models, and other statistical approaches are commonly presented in studying the variables that impact the gold price. However, these approaches have limited forecast accuracy due to the small amount of data employed. The statistical models are more suitable for a linear connection between historical data, which is impractical (Baguda and Al-Jahdali, 2021).

Time-series forecasting technique is one of the traditional machine learning models extensively used to predict finance, stock market and banking. Examples of time series forecasting techniques are the Autoregressive Integrated Moving Average model (ARIMA) and Generalised Autoregressive Conditional Heteroskedasticity (GARCH). However, this technique may not be suitable for predicting gold prices due to its weakness of being computationally expensive, unable to process massive data, and poor performance for long-term forecasts (Bora, 2021).

With the advent of machine and deep learning models, the issue of forecasting non-linear historical data may be greatly assisted with high accuracy. Machine learning models such as M5P model tree (Patalay and Bandlamudi, 2021) is able to yield a high forecasting accuracy of 85%, while deep learning model such as LSTM (Livieris, Pintelas and Pintelas, 2020; Mohtasham Khani, Vahidnia and Abbasi, 2021) is also able to identify short-term and long-term dependencies with high accuracy.

Numerous studies have been conducted to assess the impact of macroeconomic factors, stock indexes, and commodities on gold prices. Furthermore, these factors are frequently leveraged to enhance the accuracy of gold price forecasts. Nevertheless, there is limited research that combines all three elements in a single prediction system, and even fewer studies have explored the influence of copper prices on gold prices.

Several studies support the viability of technical analysis to predict commodities prices (Wang, Liu and Wu, 2020; Narayan, Ahmed and Narayan, 2015), oil prices (Yin and Yang, 2016) and stock prices (Zhai, Hsu and Halgamuge, 2007). However, technical analysis has seldom been used in predicting gold price compared to macroeconomics, stock indexes and commodities factors. Moreover, the effect of incorporating technical analysis indicators with economic factors as a predictor is seldom investigated.

Plenty of research uses time series models, machine learning models, deep learning models or hybrid models to forecast gold prices. There is also plenty of research comparing the performance of different models. However, there is a lack of study between the difference in performance for machine learning and deep learning models.

To overcome the limitation and fill the gap mentioned above, this study includes the following elements: (1) comparing and analysing the performance of forecasting the gold price of machine learning model M5P and deep learning model LSTM to determine their suitability in predicting gold price, (2) a comprehensive independent feature for gold price prediction consist of economics factors such as US dollar index, interest rates, inflation rates, US treasury bond yield, S&P500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil price were considered, (3) technical indicators of Bollinger Band, SMA, MACD, RSI, and ROC to examine its the effectiveness in gold price prediction.

1.3 Research Questions

1. What is the relationship between the US dollar index, interest rates, inflation rates, US treasury bond yield, S&P500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil price with the price of gold?
2. Which of the proposed prediction models (M5P model tree or LSTM) better accurately forecasts the gold price?
3. Will the use of technical analysis as predictor helps to increase the prediction performance of the prediction models?

1.4 Research Objectives

1. To perform analysis to examine the relationship of the US dollar index, interest rates, inflation rates, US treasury bond yield, SP500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil price on the price of gold.
2. To determine and compare the prediction performance of the M5P model tree and LSTM on the gold price.
3. To determine and compare the performance of the models with predictors of technical analysis, economic variables and a combination of both.

1.5 Research Significant and Motivation

For government officials and investors in nations whose economies are reliant on the gold commodity, predicting changes in the gold price has become a vital consideration in the current global economy. Risk and profit potential are integral components of investing in financial markets, particularly with regard to gold. Developing methods to accurately forecast gold prices is essential for investors, banks, governments, and stock markets to maximise investment returns. More precisely, understanding how various factors influence gold prices will provide critical insight and assistance for investors, financial institutions, and researchers seeking to produce efficient strategies and plans to adjust their investment portfolios amid different financial market conditions to minimise risk. Additionally, this research is significant in uncovering the efficacy of using technical analysis within machine learning to predict gold prices.

CHAPTER 2: LITERATURE REVIEW

In this study, the relationship between the US dollar index, interest rates, inflation rates, US treasury bond yield, SP500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil price with the price of gold will be determined. The M5P model tree and LSTM machine learning model will be used to predict the gold price. To gain more understanding and familiarity with current research, several past studies deal with the price of gold existing researchers have done.

2.1 Prior Research for Features Used in Predicting the Price of Gold

Numerous factors affect the price of gold, but the primary ones that are frequently studied and will be reviewed in literatures in the following section for their effects on gold include the US dollar index, interest rates, inflation rates, and crude oil prices.

2.1.1 Macroeconomic variable

The US dollar index is the primary measure used to track changes in the US dollar's value relative to a range of foreign currencies. Gold price is determined on the price scale of the US dollar. As a result, when the US dollar index falls, it indicates that the value of the US dollar has declined, meaning that the price of gold has risen. When the dollar index rises, it shows that the value of the US dollar has increased, implying that the price of gold has decreased (Qian et al., 2019). However, Arfaoui & Ben Rejeb (2017) stated that there is the possibility that both the USD and the gold price will rise at the same time. This situation might happen if a crisis occurred in another country or region that drives investors to migrate to safer assets such as gold and the US dollar. The study's findings also reveal that oil prices and the USD favorably and considerably impact gold. On the other hand, the multiple regression analysis results from Laily et al. (2017) show no significant relationship between the gold price and the USD exchange rate, which indicates that even if the exchange rate rises by one dollar, the gold price will remain unchanged. Furthermore, the study's correlation of coefficient analysis reveals that the USD exchange rate has a poor correlation with the gold price.

The interest rate is one of the measures governments take to regulate and control the country's economy. The alteration of interest rates can change the country's supply and demand for money, impacting macroeconomics (Qian et al., 2019). Typically, interest rates have an inverse effect on the gold price. When the interest rate increases, the return from the individual's savings will increase, eliminating the need for investors to invest in the gold market. However, when the interest rate lowers, investors will have more desire to invest in gold, which drives the price of gold. The relationship between interest rates and gold prices can be proved by Kan & Serin (2022), exploring the dynamic relationships between gold prices and selected financial indicators, including interest rates in Turkey from 2000 until 2019. The Dynamic Least Squares (DOLS) estimation method was used to estimate long-term coefficient results. The results found that gold prices drop by 0.48 percent for every one percent increase in interest rates. This is further supported by Lee et al. (2021) using linear regression, which found that there is only a significant level of 1% which indicated that the probability of gold price inversely related to the interest rates is very high before and during the Covid-19 period. Aside from that, Md. Isa et al. (2020) also discovered that interest rates negatively influence the gold price. The study concluded that investors would gravitate toward debt-based assets rather than gold ones when interest rates rise. Moreover, the cost of funds to invest in gold increases during high-interest rates, lowering investors' desirability to borrow for gold investment.

Inflation is defined as an increase in overall price levels. It is regarded as a macroeconomic issue since it has the ability to reduce people's purchasing power. The inflation rate is one of the potential factors that largely influence the gold price, where gold prices tend to increase during times of high inflation. According to Liya et al. (2021), the results of correlation analysis show a correlation of 0.1032 between inflation rates and gold prices, indicating a positive relationship. Moreover, the linear regression study shows a significant relationship between inflation rates and gold prices with a probability value is 0.003, which is lower than the significant level of 1%. Aside from that, the findings of regression analysis from Manjula & Karthikeyan (2019) stated that monthly gold price has a strong positive correlation with inflation rates where the correlation coefficient value is 0.94. This is further supported by Md Isa et al. (2020) with the results from the Pearson correlation test with correlation coefficient values of 0.856. The study concludes that gold works as a hedging tool in the event of inflation, where general goods prices rise. The growth of the value of gold is believed to be able to offset the rising cost of products and services. Hence, the increase in the inflation rate would result in a rise in the gold price and vice versa.

2.1.2 Commodities prices

Crude oil prices are closely related to inflation. Hence inflation is quite likely to occur when crude oil prices increase. Paper currencies will lose purchasing value over time during inflation, but gold doesn't. As a result, during the increased crude oil price, gold is favourable to be invested in large quantities to protect from inflation where other currencies lose value (Dr. Sindhu, 2013). Kalsum et al. (2021) investigate the impact of inflation, interest rates, and global oil prices on gold price changes in Indonesia, using the US Dollar exchange rate as an intermediate variable. The author used path analysis to test the hypothesis, and the result demonstrates that the price of global oil has a direct positive influence of 0.203 on the price of gold. This implied that gold prices would also increase if international oil prices rose. This is further supported by Torki et al. (2021), which explored the impact of macroeconomic factors and oil prices and changes in each of these variables on the future gold market in Iran from 2009 to 2017 using the GARCH MIDAS Model. As a result, the oil price has a favourable impact on the volatility of gold coin futures, as it is predicted that following a 1% increase in oil prices this month, gold futures will increase by 1.4% next month. Aside from that, Singhal et al. (2019) performed ARDL Bound testing cointegration technique to investigate the dynamic link between international oil prices, the stock market index, the exchange rate, and the international gold prices in Mexico. According to the study, there is a positive association between oil prices and gold prices, with a correlation coefficient of 0.241.

Research from Liu et al. (2017) investigated Pearson's correlation of crude oil, natural gas, gold, silver, lean hogs, coffee, and the Dow Jones index with copper prices to predict the copper price. With Pearson's correlation coefficient of 0.59, the research found that the price of copper strongly correlates to the price of gold due to the market competition between copper and gold.

2.1.3 Stock Indexes

The relationship between stock index and gold price is often seen as an indicator of economic health. When the stock market rises, gold prices usually fall, as investors switch to stocks for better returns. Conversely, when the stock market falls, gold prices tend to rise as investors seek a safe haven for their money. The research from Liu & Li (2017) indicating that stock indexes such as the Dow Jones Industrial Average (DJIA) and the S&P500 are key factors in predicting gold prices. Furthermore, they imply that stock indexes typically compete with gold price suggesting a negative relationship between the two.

2.1.4 Technical Indicators

Technical analysis involves studying historical charts and data to examine trends and identify potential buying or selling opportunities by analyst, but also proved to be worked in machine learning where it is used in predicting stock market movement (Bustos and Pomares-Quimbaya, 2020), commodities prices (Wang, Liu and Wu, 2020) and oil prices

(Yin and Yang, 2016). The research utilized several prominent technical indicators such as moving average, stochastic oscillator, rate of price change, MACD, RSI, and advance decline line to achieve excellent predictive performance. Moreover, the research from Wang, Liu and Wu (2020) also demonstrates that technical indicators produce more accurate predictions of the magnitude of commodity price fluctuations than economic variables.

2.2 Prior research for Gold Price Forecasting Method

There are several methods for predicting the price of gold. However, in the following sections, studies that employed time series analysis, machine learning, and deep learning will be reviewed for their performance and suitability for the gold prediction study.

2.2.1 Time Series Analysis Technique

Time series forecasting predicts future events based on previous trends and the assumption that future movements will follow similar patterns. Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and Autoregressive Integrated Moving Average (ARIMA) models are the most extensively utilised time series analysis technique.

For example, Yang (2019) utilised the ARIMA model to forecast and analyse daily gold prices in USD. The result from the study shows high accuracy, with the relative error between the anticipated and actual values being all under 1.2 percent. However, the ARIMA model is better suited to short-term forecasting as the relative inaccuracy of the model prediction increases with the forecast period duration. Furthermore, the ARIMA model assumes that the data follows a linear pattern without considering influencing variables that might produce substantial fluctuation.

On the other hand, Hasanah et al. (2019) employ GARCH and ARIMA models to predict and analyse the gold return volatility to solve the heteroscedasticity problem in time series data. The results demonstrate that ARIMA was suitable for forecasting future returns, while GARCH was utilised to assess the volatility behaviour. Moreover, Ping et al. (2013) also employed GARCH and ARIMA models to predict Malaysia's gold bullion price. The goodness of fit measure is evaluated using Akaike's information criterion (AIC), while mean absolute percentage error (MAPE) is used to measure the forecasting accuracy. The results conclude that GARCH is more appropriate than ARIMA in forecasting Malaysia's gold bullion future values due to the MAPE and SIC values for GARCH being lower than those for ARIMA.

2.2.2 Machine learning Technique

Machine learning is a non-parametric computing technique that learns from experience and improves prediction results performance progressively during the computation process (Kim Soon et al., 2018). Unlike the time series analysis technique, machine learning algorithms are able to incorporate non-linearity by including a complex bunch of variables and can compute a massive amount of data. Various machine learning models exist, including Support Vector Machines (SVM), linear regression models and tree-based algorithms such as Random Forest, bagging, gradient boosting regression trees, stochastic gradient boosting and M5P.

Researchers have employed a range of machine learning algorithms to predict gold price and its relationship with the factors thought to impact it. For example, Manjula & Karthikeyan (2019) predicted gold price using linear regression, random forest regression, and gradient-boosting regression trees and studied selected factors influencing it. Data from 2000 to 2008 was used, then divided into two periods. By comparing the results, it is concluded that random forest regression has higher prediction accuracy for long periods, while gradient boosting regression performs well in analysing the two periods individually. Both tree-based algorithms have better prediction accuracy than linear regression.

Moreover, a study by Sadorsky (2021) used bagging, stochastic gradient boosting and random forests, and Patalay & Bandlamudi (2021) used the M5P algorithm for gold price prediction, proving the excellent performance of other tree-based algorithms. Both studies obtain excellent accuracy rates of between 85 and 90 percent. Moreover, Nwulu (2017) used decision tree-based algorithms, including M5P, Random Forest, and several other decision tree models, to perform research on predicting the price of crude oil. The results show that the M5P model has the most excellent performance. Other than that, the Support Vector Machines algorithm also proved to have high prediction accuracy. For example, Mahdi et al. (2021) were able to predict cryptocurrency returns based on gold prices using the Support Vector Machines (SVM) algorithm. The results show that SVM with radial kernel was able to get performance accuracy from 89.4 - 91.8 percent, which outperforms linear kernel with 71.8 - 74.5 percent performance accuracy and polynomial kernel with 74.8 - 79.3 percent performance accuracy.

2.2.3 Deep Learning Technique

Machine-learning algorithm is limited in the ability to process raw data or unprocessed data. Part of the data pre-processing needed for machine learning is reduced by deep learning. Deep learning algorithms can automate feature extraction, ingest and analyse text and image-based unstructured data, and reduce the need for human experts by automating feature interpretation (IBM Cloud Education, 2020). The typical machine learning models are artificial neural networks (ANN), multi-layer perceptron (MLP), convolutional neural networks (CNN) and long short-term memory (LSTM).

Researchers have employed a range of deep learning machine learning algorithms to predict the gold price. For example, Shankar & Reddy (2021) compared the ARIMA model with deep learning algorithms such as MLP, CNN, and LSTM to estimate gold prices in India using daily data from the World Gold Council from 2016 to 2020. The results show that the three deep machine learning models performed substantially better than the ARIMA model with minimal prediction error, with LSTM doing the best. Mombeini & Yazdani-Chamzini (2015) further proved the high performance of the deep learning technique, which used the ANN model to predict the gold price and compared it with the ARIMA model. The results also show that the deep learning algorithm ANN outperforms the ARIMA model on all statistical criteria, including RMSE, R2, and MAE.

2.2.4 Literature Review Comparison Table

Table 2.1 below shows the comparison table for different prediction techniques and models used by each literature for gold price prediction.

Table 2.1: Prediction techniques and models used for gold price prediction in various research.

No	Prediction Techniques	Citations	Article Titles	Predictor	Models	Results/Conclusion
1	Time Series Analysis	Yang (2019)	The Prediction of Gold Price Using ARIMA Model	Historical Gold Price	ARIMA	ARIMA model is better suited to short-term forecasting and linear data pattern without considering influencing variables.
2	Time Series Analysis	Hasanah et al. (2019)	Gold Return Volatility Modelling Using GARCH	Historical Gold Price	ARIMA and GARCH	ARIMA was suitable for forecasting future returns, while GARCH was utilised to assess the volatility behaviour.
3	Machine Learning	Manjula & Karthikeyan (2019)	Gold Price Prediction Using Ensemble-Based Machine Learning Techniques	Stock market, crude oil price, rupee dollar exchange rate, inflation and interest rate	Linear Regression, Random Forest, and Gradient Boosting Regression Trees Machine	Random forest regression has higher prediction accuracy for long periods, while gradient-boosting regression. performs well in short periods. Both tree-based algorithms have better prediction accuracy than linear regression.
4	Machine Learning	Sadorsky (2021)	Predicting Gold and Silver Price Direction Using Tree-Based Classifiers	Technical Indicators	Bagging, stochastic Gradient Boosting and Random Forests	All the models have excellent accuracy rates of between 85 and 90 percent.

Table 2.1 continued

No	Prediction Techniques	Citations	Article Titles	Predictor	Models	Results/Conclusion
5	Machine Learning	Patalay & Bandlamudi (2021)	Gold Price Prediction Using Machine Learning Model Trees	Crude oil price and S&P500	M5P	The M5P model obtain high accuracy of: R^2= 0.9588 MAE = 0.06890 RMSE= 133.14
6	Machine Learning	Nwulu (2017)	A Decision Trees Approach to Oil Price Prediction	World events impact factor, Global demand, and NYMEX future contract prices	M5P, Random Forest, Decision Stump, Random Tree and REPTree	The results show that the M5P model has the greatest performance with: CC=0.9999 MAE = 0.1135 RMSE= 0.275 RAE= 0.6662% RRSE= 1.1888%
7	Machine Learning	Mahdi et al. (2021)	A New Approach to Predicting Cryptocurrency Returns Based on the Gold Prices with Support Vector Machines During the COVID-19 Pandemic Using Sensor-Related Data	Deaths and infected cased due to COVID-19	Support Vector Machines	SVM with radial kernel is able to get performance accuracy from 89.4 - 91.8 %, which outperforms linear kernel with 71.8 - 74.5 % and polynomial kernel with 74.8 - 79.3 %.

Table 2.1 continued

No	Prediction Techniques	Citations	Article Titles	Predictor	Models	Results/Conclusion
8	Deep Learning	Shankar & Reddy (2021)	Forecasting Gold Prices in India using Time series and Deep Learning Algorithms	Historical gold prices	MLP, CNN, LSTM and ARIMA	The results show that the three deep machine learning models performed substantially better than ARIMA, with LSTM doing the best with: MAE = 0.02931 MAPE= 0.69 RMSE=0.03587
9	Deep Learning	Mombeini & Yazdani-Chamzini (2015)	Modelling Gold Price via Artificial Neural Network	Silver price, USD index, oil price, inflation rate, interest rate, stock market index and gold production	ANN and ARIMA	ANN outperforms the ARIMA model on all statistical criteria.
10	Deep Learning	Mohtasham Khani et al.(2021)	A Deep Learning-Based Method for Forecasting Gold Price with Respect to Pandemics	COVID-19 time-series data	LSTM	Best performance is achieved by using multi-variable input and 2 layers of LSTM with: RMSE= 0.0251 MSE= 0.00063 MAE= 0.01774 MSLE= 0.00018 R^2= 0.85759

2.2.5 Literature Review Conclusion

According to the literature surveyed, macroeconomic factors, commodities price and stock price all have an influence on the gold price and are useful for forecasting its future. Thus, this study has incorporated several economic factors that are known to be influential based on the literature review as well as those that may enhance the accuracy of predicting gold prices. These encompass US Dollar Index, interest rates, inflation rates, US treasury bond yield, SP500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil prices. Furthermore, technical analysis also implemented as the predictor of our model in this research as a contrast of performance different with the economic factors with the support of its effectiveness from the research (Bustos and Pomares-Quimbaya, 2020;Wang, Liu and Wu, 2020 and Yin and Yang, 2016). Several well-known technical analysis are picked which includes Bollinger Band, SMA (Simple Moving Average), MACD (Moving Average Convergence Divergence), RSI (Relative Strength Index), and ROC (Rate of Change).

Moreover, M5p and LSTM model are picked as our prediction model as these two models shows outperformed prediction performance when compared to other machine learning and deep learning model. As evidence, from table 2.1 above we can see that the LSTM model proven to be highly accurate in predicting gold prices in research done by Mohtasham Khani et al. (2021) and Shankar & Reddy (2021) while M5p also proved its performance in research done by Nwulu (2017) and Patalay & Bandlamudi (2021).

CHAPTER 3: METHODOLOGY AND WORK PLAN

3.1 Research design

This paper is quantitative research in which we examine the correlation between gold prices and economic variables such as the US Dollar Index, interest rates, inflation rates, US treasury bond yield, SP500, NASDAQ Composite Price, Dow Jones Industrial Average Price, silver, copper and oil prices. In addition, the Machine Learning model (M5p) and Deep Learning model (LSTM) are applied to forecast the 1-day and 7-day ahead gold prices. The economic data was collected from a website, while the technical indicators were derived from historical gold price computations. Ultimately, the economic variables, technical indicators and their combination act as inputs for the suggested models.

3.2 Data Science Project Framework

This research applied the OSEMN framework of data science as discussed by Kumari, Bhardwaj and Sharma (2020), which demonstrates the general workflow that data scientists typically perform in a data science project. The OSEMN framework is comprised of data gathering (Obtain), data preparing (Scrub), data exploring (Explore), data modelling (Model) and lastly, result interpreting (Interpret). Figure 3.1 show the data science process flow of the OSEMN framework. Each step of the process will be discussed in the following topics.

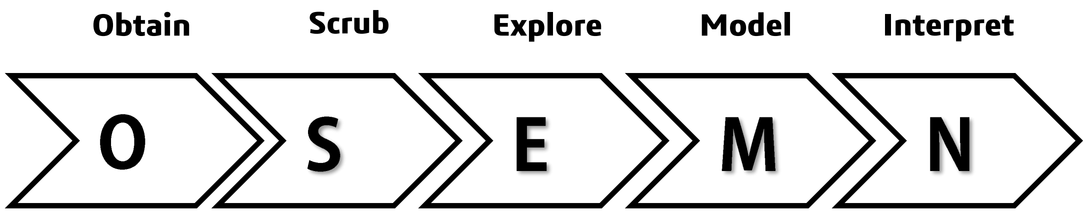

Figure 3.1: OSEMN framework of data science

3.3 Data Collection (Obtain)

3.3.1 Collected Data Source and Type

In this research, 4 commodities' historical data is collected, which includes gold (Gold Spot US Dollar), silver (Silver Futures Price), copper price (Copper Futures Price) and oil price (Brent Crude Oil Futures prices). The macroeconomics input variables were US dollar index (USDX), interest rates (Federal Funds Effective Rate/ Fed Fund Rates), inflation rates (Consumer Price Index), and US treasury bond yield (United States 10-Year Bond Yield). Lastly, stock indexes of Standard & Poor's 500 Index (SP500), NASDAQ Composite Price and Dow Jones Industrial Average Price. The data collected are daily time series data from Investing.com from January 2002 to January 2022, all of which were sourced from the United States and priced in United States Dollars (USD). This dataset included opening prices, closing prices, daily highs, daily lows and volumes for the various variables. The closing prices of these variables were selected as input for this study, with the closing gold price as the predicted output.

3.3.2 Technical Indicators

In this research, five types of technical indicators are used as attributes to predict the gold prices, which include Bollinger Band, SMA (Simple Moving Average), MACD (Moving Average Convergence Divergence), RSI (Relative Strength Index), and ROC (Rate of Change). The technical indicators data are obtained by calculation applied to the gold price data, which are essential to provide indicators of buying or selling trends.

3.3.2.1 Bollinger Band

The Bollinger Band is an established technical analysis tool that utilizes two trendlines, situated two standard deviations away from the security's price's simple moving average (SMA), to signify when the price is excessively bought or sold. Bollinger Bands are composed of three lines: a 20-day SMA as the middle band and an upper and lower band

(Hayes, 2022). The upper and lower band are typically established two standard deviations above and below the SMA. This results in the observed price movement typically occurring within the boundaries of these bands 95% of the time.

Bollinger Bands is used to indicate the expected range in which the gold price would fall within. The graph in figure 3.2 depicts the daily movement of gold prices in USD from January 2021 to January 2022, with an upper and lower band and a 20-day Simple Moving Average (SMA) as the middle line. It has been observed that approximately 90% of price movement occurs between these two bands, thereby indicating that when prices approach the top band, it may signify an overbought market and when they move closer to the bottom band, it could suggest an oversold market.

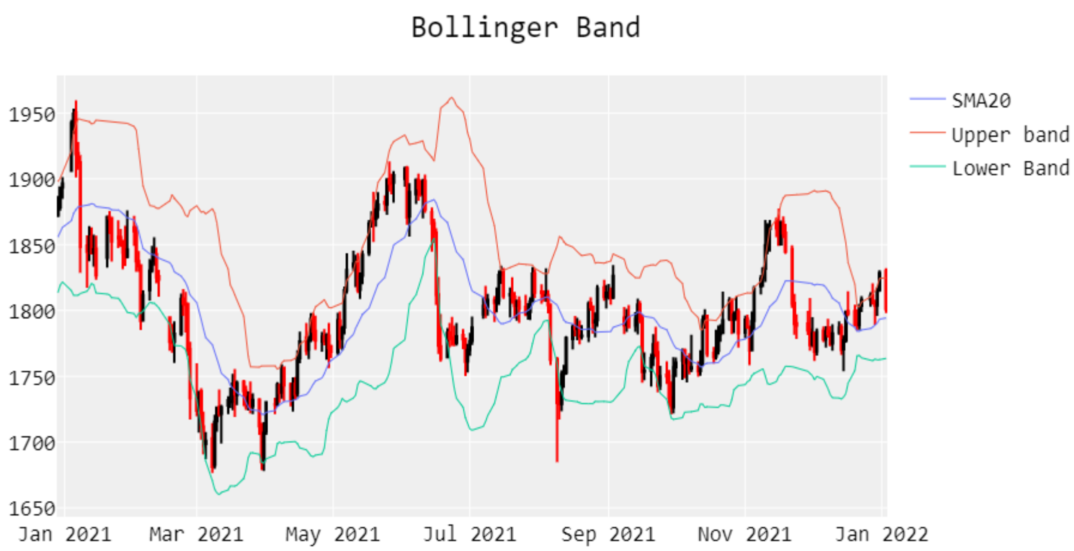

Figure 3.2: Bollinger band indicator plotted with the daily movement of gold prices in USD from period of Jan 2021 to Jan 2022

3.3.2.2 Simple Moving Average (50 Days SMA and 200 Days SMA)

Moving Average (MA) is a technical analysis tool employed to create an average price that is continuously updated to reduce the amount of noise in the price data. The angle of the moving average can then be observed, with an upward angle signifying increasing prices and a downward signifying decreasing prices, while a sideways angle generally indicates that the price is in a range (Mitchell, 2022). A Simple Moving Average (SMA) is calculated by taking the sum of past closing prices over a designated period and dividing it by the number of prices employed in the computation. This research uses both 50-day and 200-day SMAs, as denoted by Equation 3,1 below, which exemplifies a 50-days SMA taking the last fifty closing prices and dividing them by fifty.

$$SMA = \frac{\sum_{k=1}^{n} P_{t-k}}{n} \tag{3.1}$$

$$where\ n = number\ of\ days$$

The graph below in figure 3.3 displays the daily movements of gold prices in USD between January 2021 and January 2022, alongside a 200-day and 50-day SMA. The utilisation of the 50-day and 200-day SMAs enables the prediction of the gold price's directional trend. A golden cross is formed when the 50-day SMA crosses above the longer-term MA, indicating an upward trend. The opposite phenomenon, when the shorter-term MA crosses below the longer-term MA and signals a downward trend, is referred to as a "dead/death cross". Additionally, a histogram generated by subtracting 200 days SMA from 50 days SMA provides insight into both trends in the gold price and buy/sell signals from the crossover of two lines.

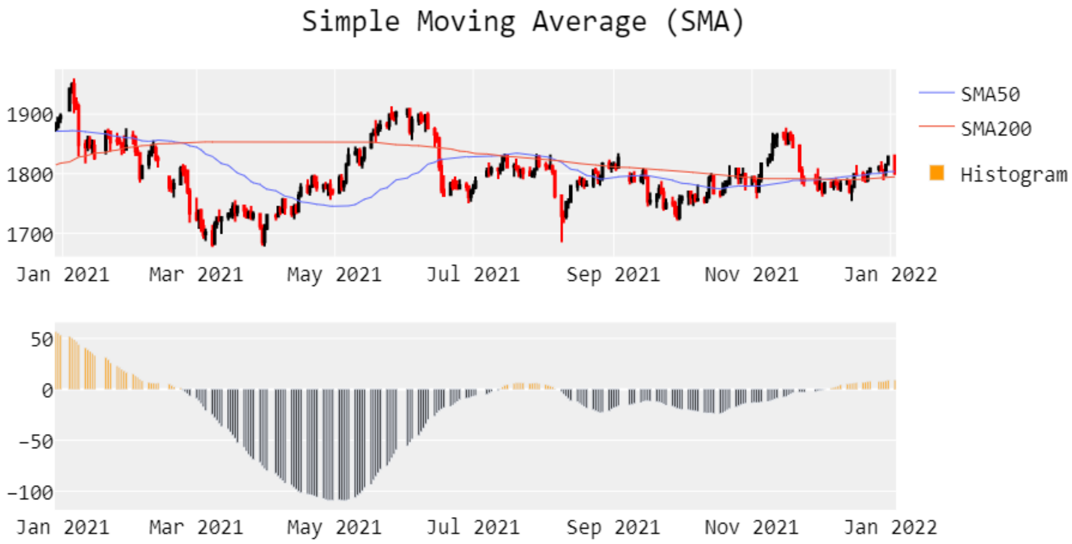

Figure 3.3: Simple moving averages indicators plotted with the daily movement of gold prices in USD from period of Jan 2021 to Jan 2022

3.3.2.3 Moving Average Convergence/Divergence (MACD)

The Moving Average Convergence/Divergence (MACD) indicator is a momentum-based indicator used to assess the momentum of a security. This is done by subtracting the 12-period Exponential Moving Average (EMA) from the 26-period EMA:

$$MACD = 26days\ EMA - 12days\ EMA \tag{3.2}$$

Where EMA is a technical indicator showing how security prices fluctuate over time, the EMA differs from a simple moving average in emphasizing recent data prices. EMA has the formula (Maverick, 2022):

$$EMA_t = \frac{2}{N+1}(P_t - EMA_{t-1}) + EMA_{t-1} \tag{3.3}$$

At recent day t, where t-1 is the previous day, and N is the number of days in the EMA, a nine-day SMA can be used as a "signal line". A buy signal is triggered when the MACD line intersects its signal line from below, while a sell signal is triggered when the MACD crosses its signal line from above.

The graph denoted as Figure 3.4 displays a graphical representation of the MACD line and nine-day SMA signal line of the gold price from January 2021 to January 2022. Additionally, a histogram is formulated by subtracting the signal line from the MACD line, which provides insight into the trajectory of the gold price as well as buy/sell signals generated from crossings between the two lines.

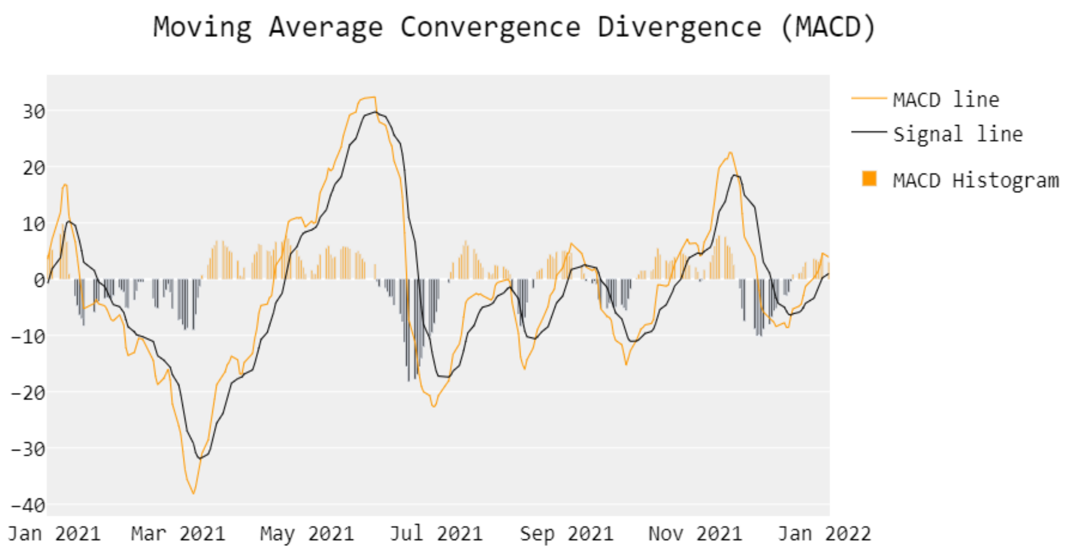

Figure 3.4: Moving Average Convergence Divergence (MACD) indicators plotted for the daily movement of gold prices in USD from period of Jan 2021 to Jan 2022

3.3.2.4 Relative Strength Index (RSI)

The relative strength index (RSI) is a momentum based indicator that measures the magnitude of price fluctuations to assess whether the asset is overvalued or undervalued, utilizing an oscillator scale ranging from 0 to 100 (Fernando, 2022). Typically, an RSI value of 70 or higher suggests an overbought market condition, whereas a reading of 30 or less indicates an oversold situation. Overbought means that a security is becoming overpriced or overly valued, while oversold means an undervalued or oversold condition where a trend reversal might happen. 14- days RSI is used in this research will have the formula:

$$RSI_{step\ 1} = 100 - \frac{100}{1 + \frac{average\ gain}{average\ loss}} \tag{3.4}$$

The average gain is the percentage of higher closes over the last 14 days, while the average loss is the percentage of lower closes in the previous 14 days. An RSI indicator of the gold price is calculated and plotted for the period of Jan 2021 to Jan 2022 in the graph above in figure 3.5 below.

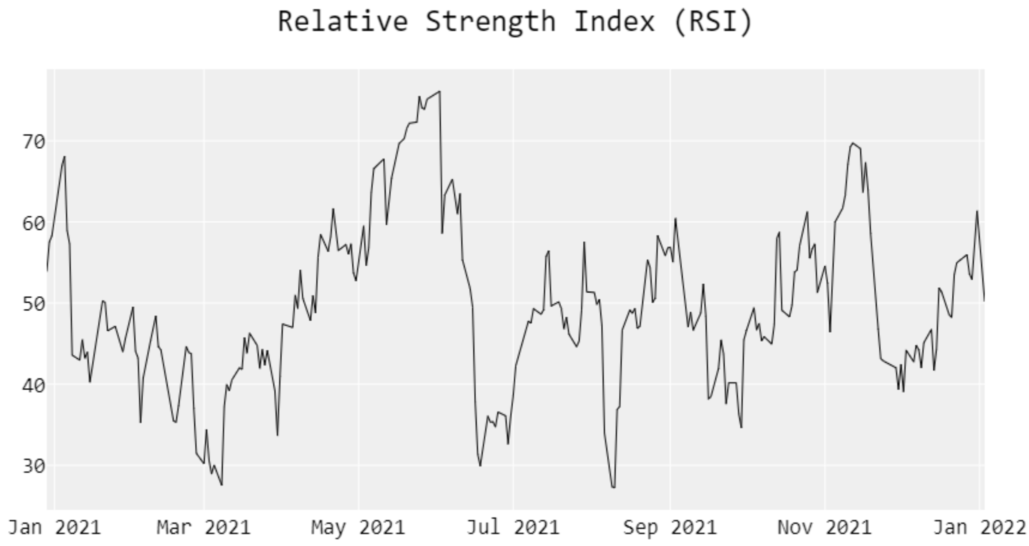

Figure 3.5: Relative Strength Index (RSI) indicators plotted for the daily movement of gold prices in USD from period of Jan 2021 to Jan 2022

3.3.2.5 Rate of Change (ROC)

The Rate of Change (ROC) is a technical indicator that measures the rate of price fluctuations, based on momentum. This indicator are able to identify divergences, overbought and oversold conditions, and centerline crossovers (Mitchell, 2021). The ROC oscillator is a technical analysis indicator with an unscaled line set against a zero-level midpoint. A rising ROC over zero usually indicates an uptrend, whereas a decreasing ROC below zero usually suggests a downtrend. When the market consolidates, the ROC will be close to zero. A 14-day ROC is derived by taking the initial price at a given time minus the price of gold 14 days ago and dividing that result by the initial price:

$$ROC = \frac{P_t - P_{t-14}}{P_t} \times 100 \qquad (3.5)$$

Where P_t is the recent closing price and P_{t-14} is the closing price 14 days ago. A ROC indicator of the gold price is calculated and plotted for the period of Jan 2021 to Jan 2022 in the graph above in figure 3.6 below.

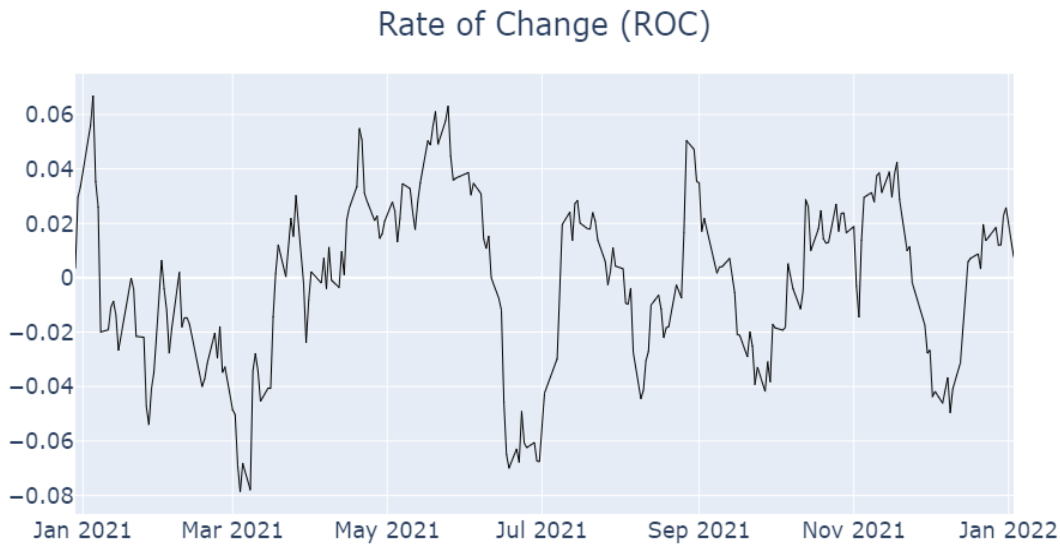

Figure 3.6: Rate of Change (ROC) indicators plotted for the daily movement of gold prices in USD from period of Jan 2021 to Jan 2022

3.4 Data Pre-processing (Scrub)

Data pre-processing steps were done to ensure the data could be more easily and effectively processed into the model for gold price forecasting. There were 5210 samples in total before data pre-processing which ranged from January 2002 until January 2022. The Consumer Price Index factor, which is monthly based, is translated to daily based by filling the current month values to every day of the month. All the collected data were merged in a data frame by matching the dates of the attributes with the target gold price.

The data found missing values due to closed markets during public holidays or major events. The rows with missing values are deleted before calculating and merging the technical analysis data with the collected data. Calculating technical analysis necessitates past records of gold prices; for instance, a SMA 200 technical analysis requires 200 previous daily data points of gold prices to generate. Therefore, after computing and combining the technical analysis with the data, the initial 200 records of the united data will be removed since the rows will feature null values.

In order to predict the gold price for one day and seven days in advance, two lagged gold prices are created by shifting one day and seven days back with respect to the attributes; the final seven rows of the dataset are then excluded due to the blank value that occurred after shifting the gold price. Subsequently, the combined dataset is split into target and features columns. The target variable is the shifted gold price, and the features are divided into three categories: macroeconomic variables, technical indicators, and all the attributes. After data pre-processing, the total sample size was 4794, spanning from 25th October 2002 until 31st December 2021. In the next step, the target variable and the feature variables are standardised using normalisation. Normalisation is a transformation technique where values are shifted and reorganised so that they span from 0 to 1. It is also referred to as Min-Max scaling.

Before we fit the input characteristics variables and output goal variables to our model, we do a sequential train-test split process to divide the input and output into 80% training data and 20% testing data. The training data is used to fit into our model to train the model while the testing data is used to test how our model would perform with new data. The training data includes information from 25th October 2002 to 28th February 2018, while the testing data consists of information from 28th February 2018 to 31st December 2021.

3.5 Exploratory Data Analysis (Explore)

3.5.1 Gold Price Data Information

The target of prediction in this study is the daily gold prices in USD from January 2002 to January 2022. Fig. 3.7 below displays a graphical representation of the daily gold prices of the stated period.

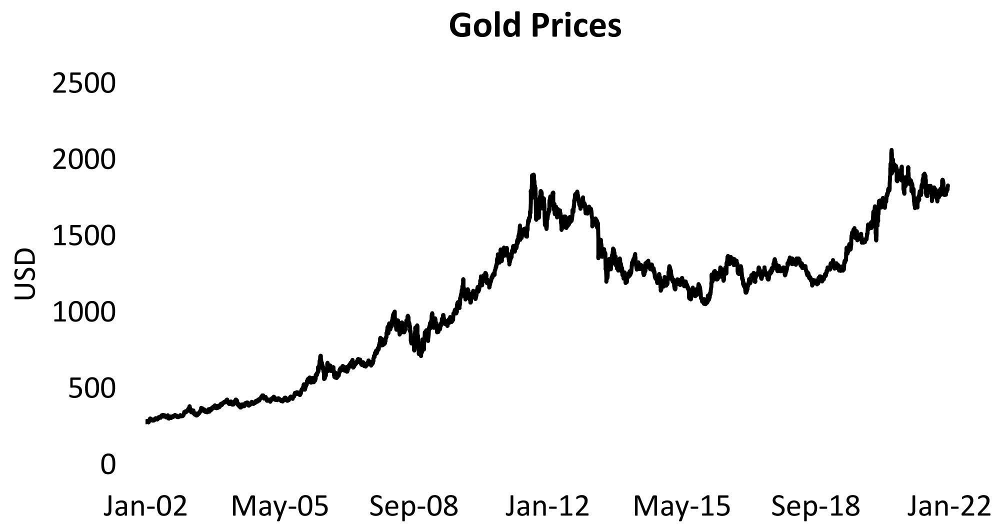

Figure 3.7: Daily gold price from January 2002 to January 2022

Table 3.1 offers descriptive metrics such as the minimum, mean, maximum, median, and standard deviation from January 2002 to January 2022 to depict the characteristics of the distribution.

Table 3.1: Statistical information of gold price data collected from January 2002 to January 2022.

Number of gold price observations	5210
Mean closing price of gold	1086.72
Standard deviation of closing price of gold	472.06
Minimum closing price of gold	278.45
25% closing price of gold	648.26
50% closing price of gold	1208.99
75% closing price of gold	1381.90
Maximum closing price of gold	2063.81

3.5.2 Correlation analysis

We performed a correlation analysis to establish which of the nine factors we gathered are strongly connected to the gold rate. Figure 3.8 illustrates the heatmap of the correlation between the variables. The correlation analysis result of the gold price with the variables was compiled and arranged in the graph shown in figure 3.9.

This correlation study reveals numerous interesting points. The factors that have a positive relationship with gold prices show that the variables move in the same direction. The most noticeably positive association is with US inflation rates (Consumer Price Index). Silver and copper prices also look closely linked to gold prices, confirming that copper correlates substantially with gold. Stock values such as NASDAQ, DJI, and S&P500 indexes also appear to have a similar impactful correlation with gold prices, with a correlation index of 0.65 to 0.67. The least positively correlated variable is Brent Crude Oil prices, but it still has a 0.5 correlation value.

On the flip side, three variables have an inverse relationship: the US dollar index, interest rates (Federal Funds Effective Rate/ Fed Fund Rates), and US treasury bond yield (United States 10-Year Bond Yield), which are macroeconomic factors. The US treasury bond yield has the most significant correlation with the gold price, while the US dollar index has a minimal correlation.

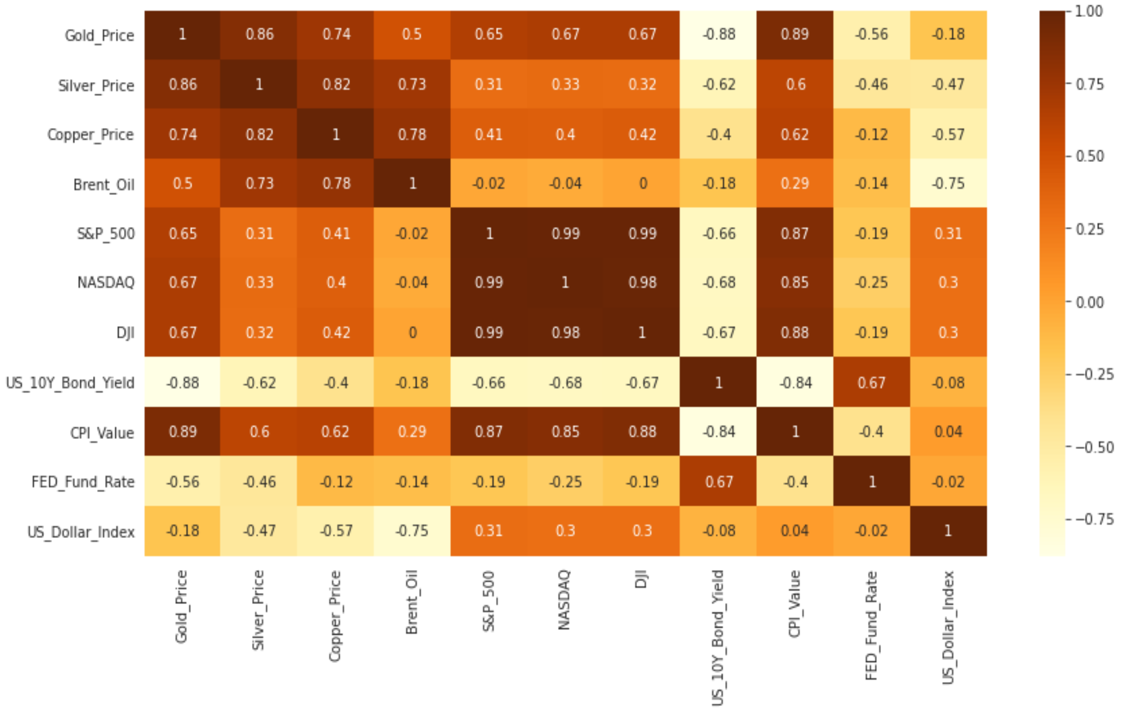

Figure 3.8: Heatmap of correlation analysis of gold prices and the other factors

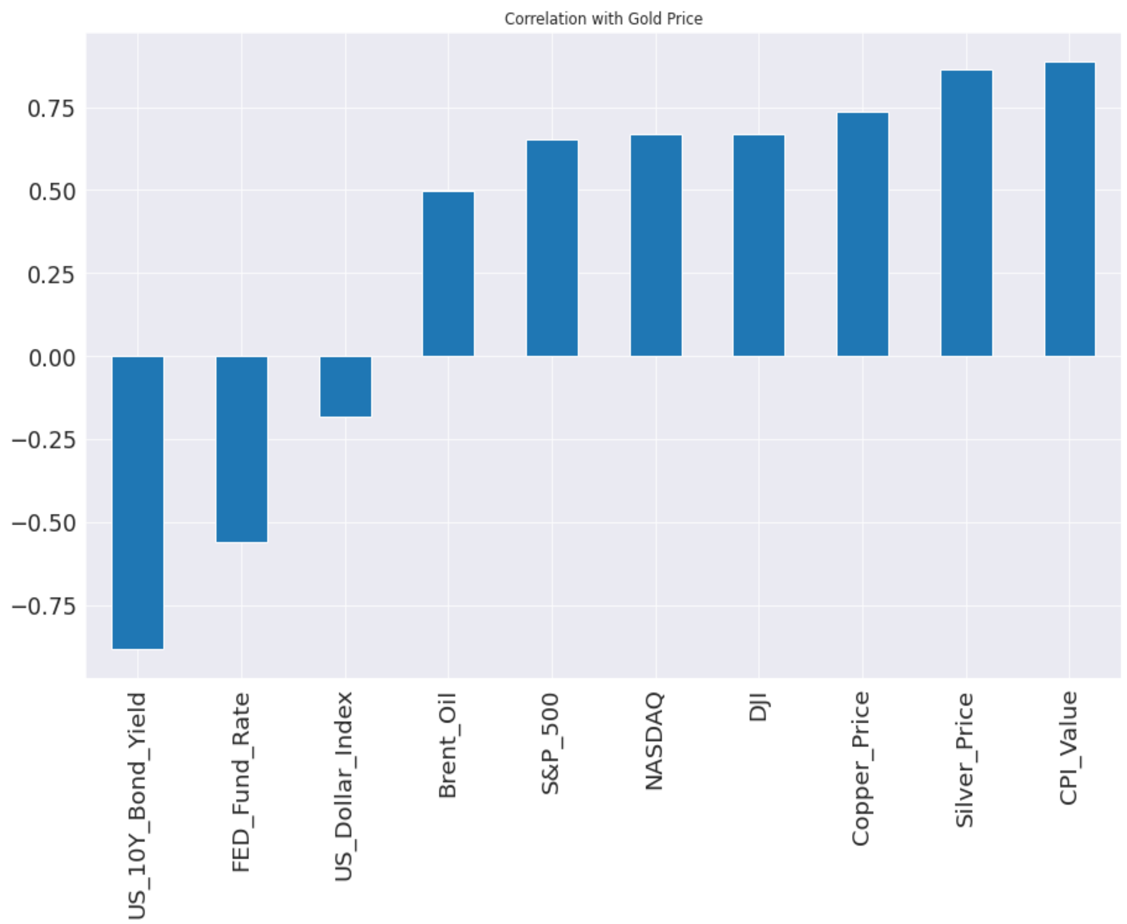

Figure 3.9: Graph of correlation analysis between gold prices and the other factors.

3.6 Model Training (Modelling)

3.6.1 Long Short-Term Memory Network (LSTM)

The Long Short-Term Memory Network (LSTM) is a enhanced type of Recurrent Neural Network (RNN) which incorporates persistent memory operations. LSTM is a recurrent neural network that facilitates the storage of temporal data (Brownlee, 2017). It can address the issue of the vanishing gradient problem that RNNs face. Each LSTM unit consists of a memory cell as well as three primary gates: a forget gate that remembers data across arbitrary time intervals and two additional gates called input and output gates that govern information into and out of the cell (Mohtasham Khani, Vahidnia and Abbasi, 2021). By choosing which information to "forget" and which to "remember," the LSTM uses this structure to create a regulated flow of information that enables it to avoid long-term dependencies (Livieris, Pintelas and Pintelas, 2020). LSTM consists of three gates: input gates, forget gates and output gates, shown in the architecture of LSTM figure 3.10 below.

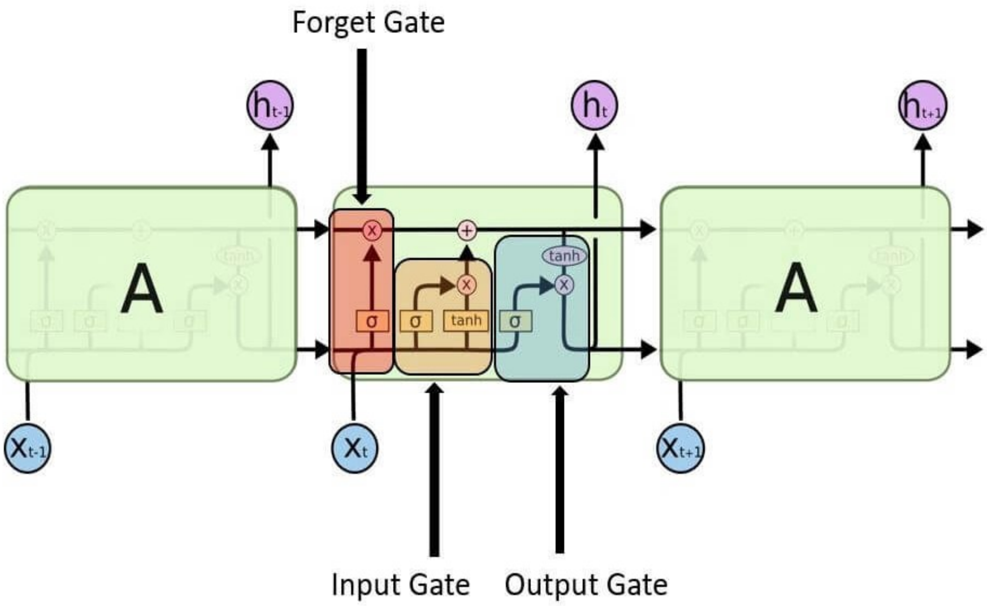

Figure 3.10: Structure of LSTM (Kalita, 2022)

The details of the gates are shown in equation 3.6-3.11, where the input gate i_t in combination with a second gate N_t, regulates the information that is retained in the memory state c_t at time t.The forget gate f_t controls whether earlier data must be deleted or kept on the memory cell at time $t-1$, whereas the output gate o_t determines which information can be utilized for the output of the memory cell (Livieris, Pintelas and Pintelas, 2020).

Input gate:

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \tag{3.6}$$

New information:

$$N_t = \tanh(U_c x_t + W_c h_{t-1} + b_c) \tag{3.7}$$

Updating cell state:

$$c_t = f_t c_{t-1} + i_t N_t \tag{3.8}$$

Forget gate:

$$f_t = \sigma(U_g x_t + W_g h_{t-1} + b_g) \tag{3.9}$$

Output gate:

$$O_t = \sigma(U_O x_t + W_O h_{t-1} + b_O) \tag{3.10}$$

Where x_t represent the input, W and U are weight matrices, b are the vector bias term, σ is the sigmoid function. Finally, the current hidden state h_t is calculated using O_t and tanh of the updated cell state:

$$h_t = o_t \times \tanh(c_t) \tag{3.11}$$

In our research, we execute single-layer and multi-layer LSTM to estimate gold prices a day in advance to test the effectiveness of the LSTM in the different numbers of layers. In the next sections, we will examine these approaches and provide a framework for comparison and choosing the most suitable model.

3.6.1.1 LSTM Methodology

A double and triple feedforward stacked LSTM model is developed to anticipate the gold rate one day and seven days ahead, as shown in figure 3.11 below. It has multiple linear stacks of LSTM layers to predict the gold price. A selection of input data Xn is put into the initial LSTM layer and set to return sequences. This is done to guarantee that the subsequent LSTM layer receives sequences instead of randomly distributed data. After going through all the layers, we have the last dense layer as a totally connected layer with neurons equal to one that produces the forecasted gold price result as Y. The LSTM models were educated for 100 epochs with an adaptive moment estimation (ADAM) optimiser, utilising a batch size equal to 64 and taking the mean-squared error function to calculate the loss. ADAM algorithm guarantees that the learning steps taken during training period are proportionate to the gradients of the parameter. The implementation code was composed in Python 3.8.10 on Google Colab. The deep learning models were crafted using the TenserFlow Keras library.

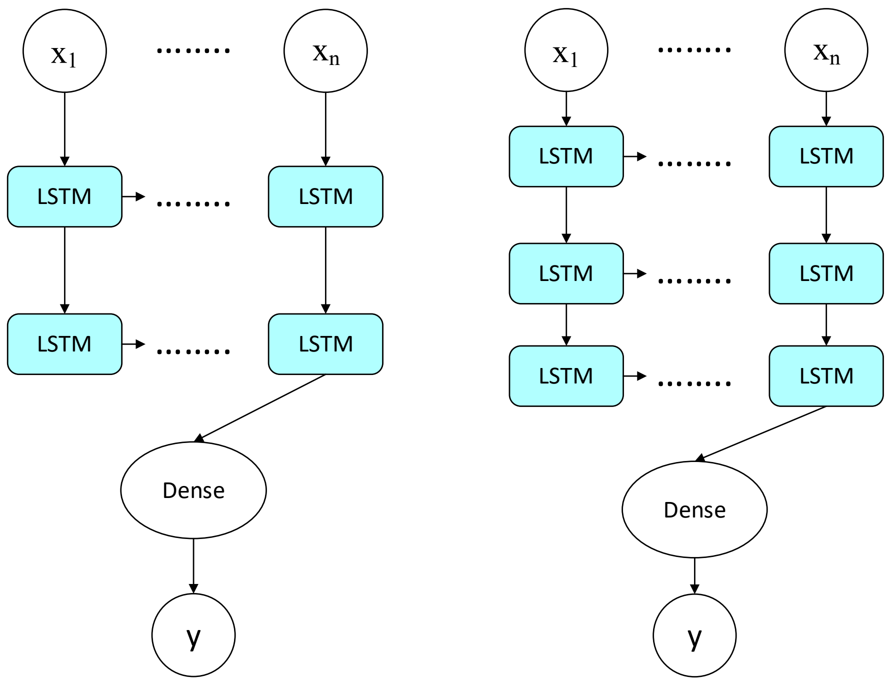

Figure 3.11: Double and triple layer sequential feedforward stacked LSTM network structure.

The Long Short-Term Memory (LSTM) layer has a variety of memory units, and a higher quantity of LSTM cells within the layer would allow for an extended memory. Thus, for longer historic days, the width of the LSTM network can be increased to achieve an optimal fit, and vice versa. The network should completely comprehend the issue without overfitting or underfitting with optimal nodes and layers to reach maximum precision. Selecting the correct number of hidden layers for a neural network is difficult, as too few layers can lead to underfitting, and too many can result in overfitting (Uzair and Jamil, 2020). Overfitting occurs when the network matches the data too closely and loses its generalisation ability. Underfitting, also known as undertraining, occurs when there are too few layers, leading to inefficient results and low time complexity. In this research, we use a different combination of 50, 100, 150 and 200 memory cells with 2 and 3 layers, as shown in table 3.2, to predict gold price one day forward and 7 days forward to determine the most effective LSTM network.

Table 3.2 : Table of LSTM Networks

Number of Layers	LSTM Network
2	L50-L50-D1
2	L100-L50-D1
2	L150-L100-D1
2	L200-L150-D1
3	L50-L50-L50-D1
3	L100-L50-L50-D1
3	L150-L100-L50-D1
3	L200-L150-L100-D1

*L denotes LSTM layers, d denotes Dense layers

3.6.2 M5p Model Tree

The M5P model tree method predicts stock prices by combining the capabilities of classification and regression. The M5p model trees are analogous to typical decision trees, however each node is fitted with a linear regression model that can forecast the class value of any case routed to it (Patalay & Bandlamudi, 2021). The tree-splitting criteria are employed to assess which attribute is most effective for partitioning the portion T of training data that arrives at a specified node. The error at a node can be calculated by calculating the standard deviation of the class values in T, and the potential decrease in error resulting from evaluating each attribute at that node can be estimated. The attribute with the greatest anticipated decrease in error is selected for division at the node.

3.6.2.1 M5p Model Tree Methodology

In order to effectively train the M5p model tree, two parameters, "max_depth" and "min_samples_leaf", must be considered when constructing the model. The former determines the maximum depth of the tree, while the latter specifies the minimum sample size needed at a leaf node. In this project, the min_sample_leaf parameter is kept at 3 as increasing this value does not significantly affect the model's performance. Conversely, when constructing the initial decision tree, maximum tree depth is a threshold that prevents additional splitting of nodes once the designated tree depth has been achieved.

By increasing the maximum depth, the predictive accuracy can be potentially improved. It has been suggested that, in general, increasing the number of splits within a decision tree can enhance predictive power so long as the model does not suffer from overfitting(Ellis, 2021). However, it is noted that the benefits of additional splits begin to diminish after a certain point. To determine the optimal number of max depths, M5p models are tested with values ranging from 2 to 5; these specifications are displayed in table 3.3 below.

Table 3.3: Specification of max_depth and min_sample_leaf value for M5p models

Model	max_depth	min_samples_leaf
1	2	3
2	3	3
3	4	3
4	5	3
5	6	3

The implementation code was composed in Python 3.8.10 on google colab. The M5p models were crafted using m5py from the Scikit-learn library. The M5p models were trained to forecast the gold price one day and seven days ahead, and the model with the highest accuracy was selected.

CHAPTER 4: RESULTS

4.1 Performance Evaluation (Interpret)

The regression prediction model's prediction accuracy is measured using Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and coefficient of determination (R^2).

4.1.1 LSTM Model

Tables 4.1 and 4.2 display the performance of the suggested LSTM networks with a different number of nodes and layers to forecast gold prices in periods of 1 and 7 days onwards respectively. From table 4.1 which the model predicts 1-day ahead gold price, we are able to obtain the highest R^2 value of 0.7440, 0.9939 and 0.9860 for economic input variables, technical indicators input, and all the attributes as inputs, respectively. Moreover, for predicting the 7-days onwards gold price, we are able to obtain the highest R^2 value of 0.5435, 0.9776 and 0.9508 for economic input variables, technical indicators input, and all the attributes as inputs respectively. We are able to get the best performance LSTM model in predicting 1- day ahead gold price with technical indicators input with MAE of 0.0081, MSE of 0.0001, RMSE of 0.0112 and R^2 of 0.9939. It is then plotted in a line graph illustrated in Figure 4.1 with the actual historical gold price in black and the predicted gold price with the best in red which we can see that the predicted gold price is very close to the actual gold price.

Table 4.1: Performance of proposed LSTM networks based on MAE, MSE, RMSE and R^2 with forecasting period of 1 day ahead.

Input Variables	Number of Layers	LSTM Network	MAE	MSE	RMSE	R^2
Economic Variables	2	L50-L50-D1	0.0728	0.0078	0.0884	0.6166
	2	**L100-L50-D1**	**0.0544**	**0.0052**	**0.0722**	**0.7440**
	2	L150-L100-D1	0.0927	0.0110	0.1049	0.4594
	2	L200-L150-D1	0.1116	0.0190	0.1378	0.0681
	3	L50-L50-L50-D1	0.1399	0.0262	0.1619	-0.2864
	3	L100-L50-L50-D1	0.1206	0.0225	-0.1500	0.1033
	3	L150-L100-L50-D1	0.1058	0.0173	0.1316	0.1506
	3	L200-L150-L100-D1	0.0629	0.0069	0.0828	0.6636
Technical Indicators	2	L50-L50-D1	0.0122	0.0002	0.0153	0.9884
	2	L100-L50-D1	0.0094	0.00015	0.0121	0.99282
	2	L150-L100-D1	0.0174	0.00038	0.01962	0.98112
	2	L200-L150-D1	0.0147	0.0003	0.01729	0.98533
	3	L50-L50-L50-D1	0.0174	0.0004	0.0207	0.9790
	3	L100-L50-L50-D1	0.0092	0.0001	0.0122	0.9927
	3	L150-L100-L50-D1	0.0124	0.0002	0.0153	0.9886
	3	**L200-L150-L100-D1**	**0.0081**	**0.0001**	**0.0112**	**0.9939**
All Attributes	2	L50-L50-D1	0.0172	0.0005	0.0233	0.9733
	2	L100-L50-D1	0.0140	0.0004	0.0189	0.9825
	2	L150-L100-D1	0.0127	0.0003	0.0175	0.9850
	2	**L200-L150-D1**	**0.0131**	**0.0003**	**0.0169**	**0.9860**
	3	L50-L50-L50-D1	0.0277	0.0011	0.0329	0.9468
	3	L100-L50-L50-D1	0.0158	0.0005	0.0227	0.9747
	3	L150-L100-L50-D1	0.0141	0.0003	0.0175	0.9850
	3	L200-L150-L100-D1	0.0148	0.0004	0.0193	0.9818

Table 4.2: Performance of proposed LSTM networks based on MAE, MSE, RMSE and R^2 with forecasting period of 7 day ahead.

Input Variables	Number of Layers	LSTM Network	MAE	MSE	RMSE	R^2
Economic Variables	2	L50-L50-D1	0.0911	0.0133	0.1155	0.3474
	2	L100-L50-D1	0.0838	0.0112	0.1058	0.4516
	2	L150-L100-D1	0.0850	0.0108	0.1041	0.4697
	2	**L200-L150-D1**	**0.0757**	**0.0093**	**0.0966**	**0.5435**
	3	L50-L50-L50-D1	0.1514	0.0365	0.1912	-0.7889
	3	L100-L50-L50-D1	0.1713	0.0479	0.2188	-1.3430
	3	L150-L100-L50-D1	0.1480	0.0340	0.1844	-0.6644
	3	L200-L150-L100-D1	0.1087	0.0186	0.1363	0.0901
Technical Indicators	2	L50-L50-D1	0.0188	0.0006	0.0253	0.9685
	2	L100-L50-D1	0.0251	0.0011	0.0338	0.9440
	2	L150-L100-D1	0.0218	0.0009	0.0295	0.9574
	2	**L200-L150-D1**	**0.0153**	**0.0005**	**0.0214**	**0.9776**
	3	L50-L50-L50-D1	0.0175	0.0005	0.0228	0.9746
	3	L100-L50-L50-D1	0.0168	0.0005	0.0228	0.9746
	3	L150-L100-L50-D1	0.0197	0.0007	0.0264	0.9659
	3	L200-L150-L100-D1	0.0183	0.0006	0.0245	0.9706
All Attributes	2	L50-L50-D1	0.0262	0.0012	0.0346	0.9415
	2	**L100-L50-D1**	**0.0229**	**0.0010**	**0.0317**	**0.9508**
	2	L150-L100-D1	0.0267	0.0012	0.0350	0.9399
	2	L200-L150-D1	0.0303	0.0016	0.0396	0.9230
	3	L50-L50-L50-D1	0.0321	0.0019	0.0433	0.9083
	3	L100-L50-L50-D1	0.0646	0.0081	0.0900	0.6032
	3	L150-L100-L50-D1	0.0421	0.0028	0.0533	0.8608
	3	L200-L150-L100-D1	0.0286	0.0014	0.0372	0.9321

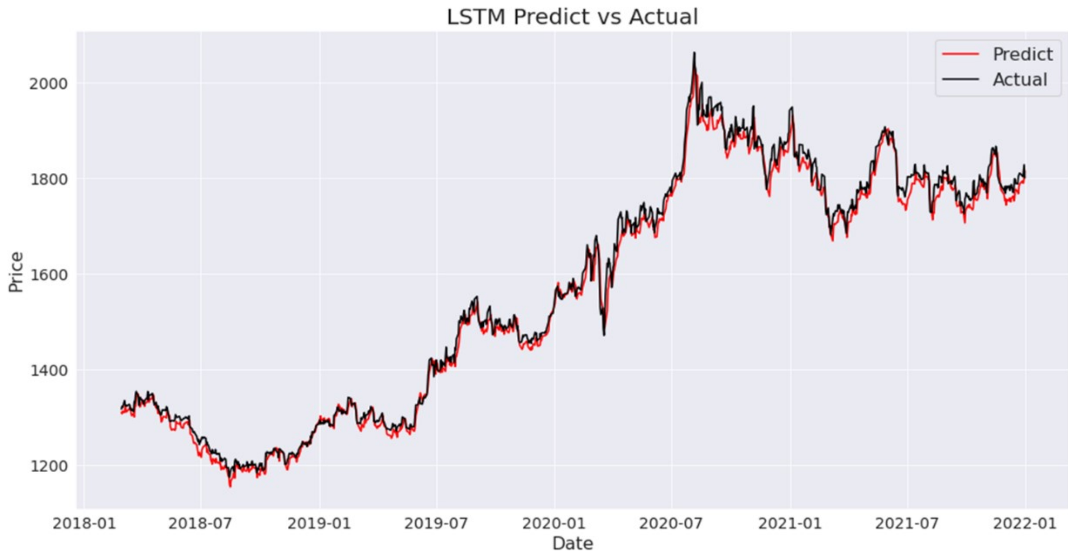

Figure 4.1: Line graph of actual historical gold price vs predicted gold price using LSTM.

4.1.2 M5p Models

Tables 4.3 and 4.4 display the performance of the suggested M5P models with a different number of max_depth settings to forecast gold prices from 1 and 7 days ahead respectively. From table 4.3 which the model predicts 1-day onward gold price, we are able to obtain the optimum values of max_depth for each model, which are 2, 6 and 6 for economic input variables, technical indicators input, and all the attributes as inputs respectively. Moreover, for predicting 7-days onwards gold price, the optimum values of max_depth for each model with economic input variables, technical indicators input, and all the attributes as inputs are 2, 4 and 4 respectively. The models mentioned above can produce the lowest MAE, MSE and RMSE values while yield the highest R^2 value. Furthermore, we are able to get the best performance M5p model in predicting 1- day ahead gold price with technical indicators input with MAE of 0.0068, MSE of 0.0001, RMSE of 0.0098 and R2 of 0.9953. It is then plotted in a line graph illustrated in Figure 4.1 with the actual historical gold price in black and the predicted gold price with the best in red which we can see that the predicted gold price is very close to the actual gold price.

Table 4.3: Performance of proposed M5p models based on MAE, MSE, RMSE and R^2 with forecasting period of 1 day ahead.

Input Variables	max_depth	min_sample leaf	MAE	MSE	RMSE	R2
Economic Variables	**2**	**3**	**0.0537**	**0.0057**	**0.0752**	**0.7226**
	3	3	0.0567	0.0065	0.0804	0.6826
	4	3	0.3683	0.1619	0.4024	-6.9428
	5	3	0.2124	0.0529	0.2299	-1.5934
	6	3	0.2335	0.0581	0.2411	-1.8522
Technical Indicators	2	3	0.0152	0.0004	0.0209	0.9785
	3	3	0.0168	0.0005	0.0231	0.9738
	4	3	0.0087	0.0002	0.0125	0.9923
	5	3	0.0069	0.0001	0.0099	0.9952
	6	**3**	**0.0068**	**0.0001**	**0.0098**	**0.9953**
All Attributes	2	3	0.0152	0.0004	0.0209	0.9785
	3	3	0.0168	0.0005	0.0231	0.9738
	4	3	0.0806	0.0124	0.1115	0.3901
	5	3	0.0070	0.0001	0.0102	0.9949
	6	**3**	**0.0070**	**0.0001**	**0.0102**	**0.9949**

Table 4.4: Performance of proposed M5p models Based on MAE, MSE, RMSE and R^2 with a forecasting period of 7 days ahead.

Input Variables	max_depth	min_sample leaf	MAE	MSE	RMSE	R2
Economic Variables	**2**	**3**	**0.0562**	**0.0064**	**0.0797**	**0.6890**
	3	3	0.2678	0.0804	0.2835	-2.9348
	4	3	0.2623	0.0756	0.2749	-2.6999
	5	3	0.2272	0.0590	0.2429	-1.8878
	6	3	0.2345	0.0610	0.2470	-1.9876
Technical Indicators	2	3	0.0249	0.0011	0.0328	0.9472
	3	3	0.0248	0.0011	0.0332	0.9459
	4	**3**	**0.0167**	**0.0005**	**0.0227**	**0.9747**
	5	3	0.0416	0.0034	0.0584	0.8330
	6	3	0.0363	0.0027	0.0515	0.8702
All Attributes	2	3	0.0249	0.0012	0.0340	0.9435
	3	3	0.0248	0.0011	0.0332	0.9459
	4	**3**	**0.0247**	**0.0010**	**0.0320**	**0.9499**
	5	3	0.0447	0.0038	0.0620	0.8119
	6	3	0.0382	0.0027	0.0520	0.8674

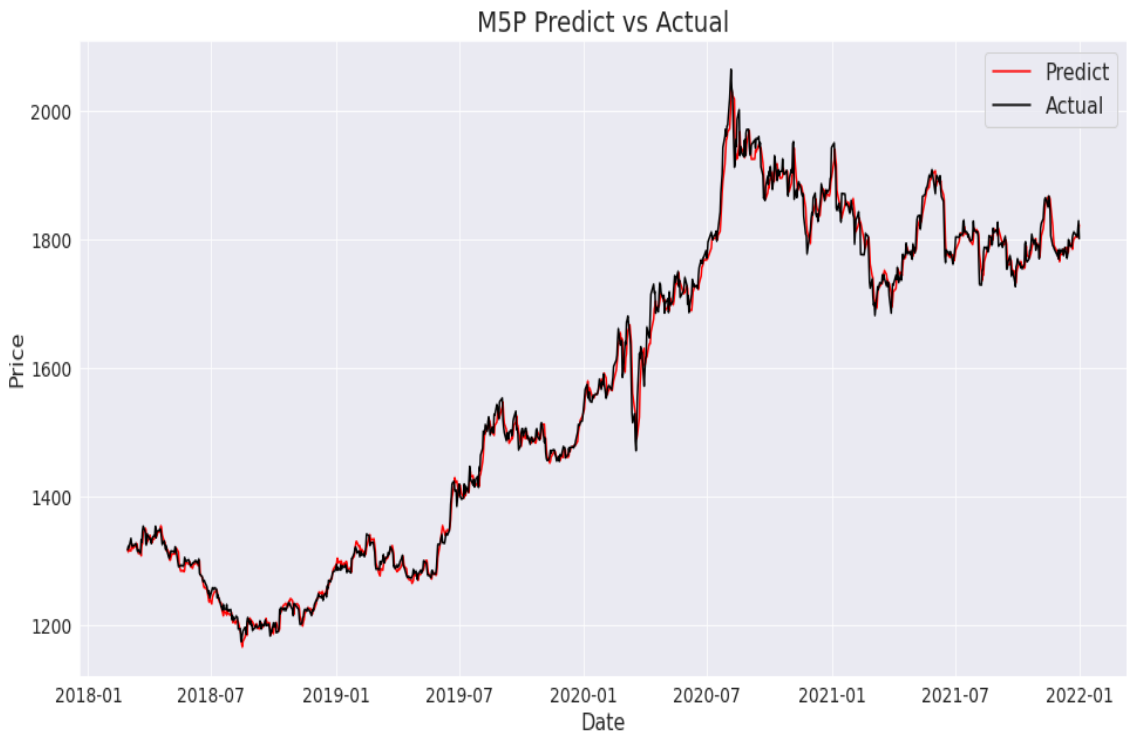

Figure 4.2: Line graph of actual historical gold price vs predicted gold price using M5p.

CHAPTER 5: DISCUSSION

5.1.1 Effect of Number of Layers and Nodes on LSTM Performance

In order to more accurately examine and discuss the impact of the number of layers and nodes on the efficacy of LSTM models, the most effective LSTM networks for each type of input and predicting period are outlined in table 5.1 below. As demonstrated in table 5.1, two layers of LSTM networks were the most effective for both prediction periods, except for 1-day ahead gold price predictions using the LSTM model with technical indicator input, where three layers produced the most optimal results. By observing the LSTM network configuration, we notice that the minimum number of nodes required in the first layers required is 100 nodes. Moreover, we observed that the optimal LSTM networks had configurations of fewer nodes on the second and third than their previous layer. This finding is in line with the principle proposed by Karsoliya (2012) that the number of neurons in the hidden layer should be approximately two-thirds of those in the input layer.

Table 5.1: Number of layers and LSTM networks configuration of the best performance LSTM models.

Input Variables	Predict Period	No of Layers	LSTM Network	MAE	MSE	RMSE	R^2
Economic Variables	1 day after	2	L100-L50-D1	0.0544	0.0052	0.0722	0.7440
	7 days after	2	L200-L150-D1	0.0757	0.0093	0.0966	0.5435
Technical Indicators	1 day after	3	L200-L150-L100-D1	0.0081	0.0001	0.0112	0.9939
	7-days after	2	L200-L150-D1	0.0153	0.0005	0.0214	0.9776
All Attributes	1 day after	2	L200-L150-D1	0.0131	0.0003	0.0169	0.9860
	7-days after	2	L100-L50-D1	0.0229	0.0010	0.0317	0.9508

5.1.2 Comparison of Model's Performance by Input Variables

Table 5.2 below shows the model with the best performance for each input variable type with a predicting period of 1 day onward and 7 days onwards for the LSTM and M5p models. The results suggest that technical analysis as input yields the highest prediction performance for both periods, with R^2 values ranging from 0.9499 to 0.9949 when using the LSTM and M5P models. However, when using only economic variables as input, the prediction performance is comparatively lower, with R^2 values ranging from 0.5435 to 0.7440. The results proved the finding of research from Wang, Liu and Wu (2020) which stated that the technical analysis perform better than economic variables. The poor performance of economics input models may be due to the fact that gold prices have risen substantially over the past few years, as well as the numerous unpredictable events and external factors which may not be taken into account by the model when analysing data from different time periods. The model has been trained on data from 25th October 2002 to 28th February 2018, spanning a period of approximately fifteen years. The testing data covers from 28th February 2018 to 31st December 2021. This difference in time frames could potentially lead to a different pattern of association between the gold price and the economic variable over the two periods. The superior performance of technical indicators-based predictive models may be attributed to the technical analysis calculated using historical gold prices, which tend not to fluctuate as much as other economic variables. Technical analysis also offers insight into the direction, volatility, range, and momentum of price movement, allowing LSTM models to forecast gold prices accurately. Furthermore, the existence of technical indicators in all attributes input yields an improvement in model performance relative to economic variable inputs; however, their performance is slightly inferior to that of technical indicator inputs alone.

Table 5.2: Performance of different input variables for LSTM and M5p models

Model	Predicting Period	Input Variables	MAE	MSE	RMSE	R^2
LSTM	1 - day ahead	Economic Variables	0.0544	0.0052	0.0722	0.7440
		Technical Indicators	0.0081	0.0001	0.0112	0.9939
		All Attributes	0.0131	0.0003	0.0169	0.9860
	7 days ahead	Economic Variables	0.0757	0.0093	0.0966	0.5435
		Technical Indicators	0.0153	0.0005	0.0214	0.9776
		All Attributes	0.0229	0.0010	0.0317	0.9508
M5P	1 - day ahead	Economic Variables	0.0537	0.0057	0.0752	0.7226
		Technical Indicators	0.0068	0.0001	0.0098	0.9953
		All Attributes	0.0070	0.0001	0.0102	0.9949
	7 days ahead	Economic Variables	0.0562	0.0064	0.0797	0.6890
		Technical Indicators	0.0167	0.0005	0.0227	0.9747
		All Attributes	0.0247	0.0010	0.0320	0.9499

5.1.3 Comparison of Model's Performance

Tables 5.3 and 5.4 display the performance of the best configurated LSTM and M5p model compared to the widely used regression models of support vector regressor (SVR) in forecasting period of 1 and 7 days ahead respectively. It is remarkable that both the LSTM and M5p models demonstrate similar admirable results and outperform the support vector regressor in all forecasting periods and input variables, attaining the lowest MAE, MSE and RMSE scores while producing the highest R^2 values. These results proved that the proposed LSTM and M5p models are well-suited for predicting gold prices in this project. We are able to get the best performance LSTM model in predicting 1- day ahead gold price with technical indicators input with MAE of 0.0081, MSE of 0.0001, RMSE of 0.0112 and R2 of 0.9939 which achieved better performance than the research done Mohtasham Khani et al. (2021) and Shankar & Reddy (2021) which able to get MAE of 0.0177 and 0.02931 respectively. Furthermore, the M5p model we are able to get the best performance M5p model in predicting 1- day ahead gold price with technical indicators input with MAE of 0.0068, MSE of 0.0001, RMSE of 0.0098 and R2 of 0.9953 which are better than the research done by Patalay & Bandlamudi (2021) and Nwulu (2017) which yield R^2 of 0.9588 and MAE of 0.1135 respectively. Moreover, it appears that for short-term gold price forecasting (1 day ahead), the LSTM model produces slightly more accurate results than LSTM when incorporating economic variable as input, while LSTM performs better with 7-days ahead prediction.

Table 5.3: Performance of LSTM and M5p model against traditional regression models based on MAE, MSE, RMSE and R^2 with forecasting period of 1 day ahead.

Input Variables	Prediction Model	MAE	MSE	RMSE	R2
Economic Variables	LSTM	0.0544	0.0052	0.0722	0.7440
	M5P	0.0537	0.0057	0.0752	0.7226
	SVR	0.0658	0.00708	0.084148	0.65266
Technical Indicators	LSTM	0.0081	0.0001	0.0112	0.9939
	M5P	0.0068	0.0001	0.0098	0.9953
	SVR	0.04557	0.00292	0.054032	0.85679
All Attributes	LSTM	0.0131	0.0003	0.0169	0.9860
	M5P	0.0070	0.0001	0.0102	0.9949
	SVR	0.09309	0.01018	0.100889	0.50071

Table 5.4: Performance of LSTM And M5p model against traditional regression models based on MAE, MSE, RMSE and R^2 with forecasting period of 7 day ahead.

Input Variables	Prediction Model	MAE	MSE	RMSE	R^2
Economic Variables	LSTM	0.0757	0.0093	0.0966	0.5435
	M5P	0.0562	0.0064	0.0797	0.6890
	SVR	0.08467	0.00838	0.09154	0.53975
Technical Indicators	LSTM	0.0153	0.0005	0.0214	0.9776
	M5P	0.0167	0.0005	0.0227	0.9747
	SVR	0.02067	0.00078	0.028004	0.96161
All Attributes	LSTM	0.0229	0.0010	0.0317	0.9508
	M5P	0.0247	0.0010	0.0320	0.9499
	SVR	0.07682	0.00698	0.083574	0.65805

CHAPTER 6: CONCLUSION

In this study, our objective to examine the correlation of the economic variables with gold prices are fulfilled by finding US inflation rates, silver prices, copper prices, NASDAQ, DJI, and S&P500 indexes are strongly positively correlated with the gold price. The least positively correlated variable was Brent Crude Oil prices; however, it still displayed a correlation coefficient of 0.5. On the flip side, there is a clear inverse relationship between the gold price and three macroeconomic factors: the US Dollar Index, interest rates, and US Treasury Bond Yield. The US Treasury Bond Yield has the strongest correlation with gold prices, while the US Dollar Index has the weakest correlation. This project employed multiple settings and configurations to train the LSTM and M5p models, and the results revealed the importance of tuning the models to identify their most optimal settings, as different settings yield substantially varied results. The best-tuned prediction results answered the research question of this project which technical indicators are more effective predictors than economic indicators which able to help to increase the prediction performance of the prediction models. Which technical analysis able to get with R^2 values of 0.97 to 0.99 which comparable with the and R^2 values of 0.54 to 0.74 for economic variable in predicting 1-day and 7-days ahead gold prices using both LSTM and M5p models. In the objective of comparing the M5P and LSTM model we found that the performance of both the model are very similar with only slight difference when technical indicators and all the attributes are used as predictor. The results also discovered that the LSTM model is slightly more accurate for short-term gold price forecasting (1-day ahead) when incorporating economic variables inputs while LSTM performs better with 7-days ahead prediction. Moreover, when comparing the LSTM and M5p models to the traditional support vector regressor (SVR) model, both M5p and LSTM yielded superior performance across all forecasting periods and input variables which prove the suitability of the proposed M5p and LSTM model in predicting gold price.

For the research impliction, this research is important which provides high performance prediction model for predicting gold price which is critical to assist investors to make more informed decisions on when to buy and sell their gold. Moreover, by understanding the relationship between economic factors with gold price are significant for investors to manipulate their investment portfolio during different financial market situation to prevent risk. Lastly, the discovering of the potential of utilizing technical analysis within machine learning to forecast gold prices is important for other researchers in building a better prediction model.

In this research, we have found several limitations which may caused the low performance of economic factors in predicting gold price which includes the huge time frame different for testing and training dataset, some economic factors with low importance to the models which may induce noise into the prediction model and the influence of major economic events are not considered as the factors which impact may generate huge impact. For future works, improvements can be made by splitting the dataset be into several smaller segments with nearer time frame and feature selection has to be conducted to eliminate the low importance variable. Lastly, the news affect might can be incorporated to give signal about potential future price movement in case of major events occurs.

REFERENCES

Arfaoui, M. and Ben Rejeb, A. (2017) 'Oil, gold, US dollar and stock market interdependencies: a global analytical insight', *European Journal of Management and Business Economics*, 26(3), pp. 278–293. doi: 10.1108/EJMBE-10-2017-016.

Baguda, Y. S. and Al-Jahdali, H. M. (2021) 'An Intelligent Gold Price Prediction Based on Automated Machine and k-fold Cross Validation Learning', *International Journal of Computer*

Bishop, K. (2020) *These 2 charts show copper prices could fall further — and the global economy too*. Available at: https://apac.cib.natixis.com/m-a-pulse-in-apac-articles/australia-focus-on/articles/the-curious-case-of-copper-and-gold-prices (Accessed: 4 June 2022).

Blaifi, S. ali *et al.* (2018) 'M5P model tree based fast fuzzy maximum power point tracker', *Solar Energy*, 163(August 2017), pp. 405–424. doi: 10.1016/j.solener.2018.01.071.

Bora, N. (2021) *Understanding ARIMA Models for Machine Learning*. Available at: https://www.capitalone.com/tech/machine-learning/understanding-arima-models/ (Accessed: 26 May 2022).

Brownlee, J. (2017) *A Gentle Introduction to Long Short-Term Memory Networks by the Experts*. Available at: https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/ (Accessed: 22 August 2022).

Bustos, O. and Pomares-Quimbaya, A. (2020) 'Stock market movement forecast: A Systematic review', *Expert Systems with Applications*, 156, p. 113464. doi: 10.1016/j.eswa.2020.113464.

Dr. Sindhu, D. S. (2013) 'A study on impact of select factors on the price of Gold', *IOSR Journal of Business and Management*, 8(4), pp. 84–93. doi: 10.9790/487x-0848493.

Ellis, C. (2021) *Max Depth In Random Forests*. Available at: https://crunchingthedata.com/max-depth-in-random-forests/ (Accessed: 29 January 2023).

Fernando, J. (2022) *Relative Strength Index (RSI) Indicator Explained With Formula*. Available at: https://www.investopedia.com/terms/r/rsi.asp (Accessed: 5 January 2023).

Gobler, E. (2022) *The Stock Market May Be Down, But the S&P 500, Nasdaq Composite, and Dow Jones Are Still Great Investments. Here's How to Get Started*. Available at: https://time.com/nextadvisor/investing/sp500-nasdaq-dow-compared/ (Accessed: 24 May 2022).

Hasanah, P., Nasir, S. Q. and Subchan, S. (2019) 'Gold Return Volatility Modeling Using Garch', *Indonesian Journal of Mathematics Education*, 2(1), p. 20. doi: 10.31002/ijome.v2i1.1222.

Hayes, A. (2022) *Bollinger Bands®: What They Are, and What They Tell Investors*, *Investopedia*. Available at: https://www.investopedia.com/terms/b/bollingerbands.asp.

IBM Cloud Education (2020) *Deep Learning*. Available at: https://www.ibm.com/cloud/learn/deep-learning (Accessed: 10 June 2022).

Kalsum, U., Hidayat, R. and Oktaviani, S. (2021) 'The Effect of Inflation, Interest Rates and World Oil Prices on Gold Prices in Indonesia with the US Dollar Exchange Rate as an Intermediary Variable', *International Journal of Finance Research*, 2(1), pp. 1–12. doi: 10.47747/financeinvestmentderivative.v2i1.228.

Kan, E. and Serin, Z. V. (2022) 'Analysis of cointegration and causality relations between gold prices and selected financial indicators: Empirical evidence from Turkey', *International Journal of Advanced and Applied Sciences*, 9(3), pp. 1–9. doi: 10.21833/IJAAS.2022.03.001.

Kalita, D. (2022) *An Overview on Long Short Term Memory (LSTM)*. Available at: https://www.analyticsvidhya.com/blog/2022/03/an-overview-on-long-short-term-memory-lstm/ (Accessed: 15 January 2023).

Karsoliya, S. (2012) 'Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture', *International Journal of Engineering Trends and Technology*, 3(6), pp. 714–717.

Khan, P., Ranjan, P. and Kumar, S. (2021) 'Data heterogeneity mitigation in healthcare robotic systems leveraging the Nelder–Mead method', *Artificial Intelligence for Future Generation Robotics*. doi: https://doi.org/10.1016/B978-0-323-85498-6.00012-5.

Kim Soon, G. *et al.* (2018) 'A CIMB stock price prediction case study with feedforward neural network and recurrent neural network', *Journal of Telecommunication, Electronic and Computer Engineering*, 10(3–2), pp. 89–94.

Kumari, K., Bhardwaj, M. and Sharma, S. (2020) 'OSEMN Approach for Real Time Data Analysis', *International Journal of Engineering and Management Research*, 10(02), pp. 107–110. doi: 10.31033/ijemr.10.2.11.

Laily, S. *et al.* (2017) 'Macroeconomic Variables Affecting the Volatility of Gold Price', *Journal of Global Business and Social Entrepreneurship (GBSE)*, 3(5), pp. 97–106.

Lee, D. *et al.* (2021) 'Does the Factors Affecting Gold Price Changed During Covid-19 Period in Malaysia ? pp. 259–269. doi: 10.6007/IJARBSS/v11-i9/10761.

Liu, C. *et al.* (2017) 'Forecasting copper prices by decision tree learning', *Resources Policy*, 52(May), pp. 427–434. doi: 10.1016/j.resourpol.2017.05.007.

Liu, D. and Li, Z. (2017) 'Gold price forecasting and related influence factors analysis based on random forest', *Advances in Intelligent Systems and Computing*, 502, pp. 711–723. doi: 10.1007/978-981-10-1837-4_59.

Livieris, I. E., Pintelas, E. and Pintelas, P. (2020) 'A CNN–LSTM model for gold price time-series forecasting', *Neural Computing and Applications*, 32(23), pp. 17351–17360. doi: 10.1007/s00521-020-04867-x.

Liya, A. *et al.* (2021) 'How macroeconomic indicators influence gold price management', *Business Process Management Journal*, (July). doi: 10.1108/BPMJ-12-2020-0579.

Mahdi, E. *et al.* (2021) 'A new approach to predicting cryptocurrency returns based on the gold prices with support vector machines during the COVID-19 pandemic using sensor-related data', *Sensors*, 21(18). doi: 10.3390/s21186319.

Manjula, K. A. and Karthikeyan, P. (2019) 'Gold price prediction using ensemble based machine learning techniques', *Proceedings of the International Conference on Trends in Electronics and Informatics, ICOEI 2019*, 2019-April(April 2019), pp. 1360–1364. doi: 10.1109/icoei.2019.8862557.

Maverick, J. B. (2022) *How Is the Exponential Moving Average (EMA) Formula Calculated?* Available at: https://www.investopedia.com/ask/answers/122314/what-exponential-moving-average-ema-formula-and-how-ema-calculated.asp#:~:text=The exponential moving average (EMA) is a technical chart indicator,importance to recent price data. (Accessed: 5 January 2023).

Md Isa, M. A. *et al.* (2020) 'Relational Study Between Macroeconomic Variables and Gold Price: Latest Malaysian Evidence', *Advanced International Journal of Business, Entrepreneurship and SMEs*, 2(6), pp. 01–09. doi: 10.35631/aijbes.26001.

Mitchell, C. (2021) *Price Rate of Change (ROC) Indicator: Definition and Formula*. Available at: https://www.investopedia.com/terms/p/pricerateofchange.asp (Accessed: 5 January 2023).

Mitchell, C. (2022) *How to Use a Moving Average to Buy Stocks*. Available at: https://www.investopedia.com/articles/active-trading/052014/how-use-moving-average-buy-stocks.asp#toc-why-use-a-moving-average (Accessed: 5 January 2023).

Mohtasham Khani, M., Vahidnia, S. and Abbasi, A. (2021) 'A Deep Learning-Based Method for Forecasting Gold Price with Respect to Pandemics', *SN Computer Science*, 2(4), pp. 1–12. doi: 10.1007/s42979-021-00724-3.

Mombeini, H. and Yazdani-Chamzini, A. (2015) 'Modeling Gold Price via Artificial Neural Network', *Journal of Economics, Business and Management*, 3(7), pp. 699–703. doi: 10.7763/joebm.2015.v3.269.

Narayan, P. K., Ahmed, H. A. and Narayan, S. (2015) 'Do Momentum-Based Trading Strategies Work in the Commodity Futures Markets?', *Journal of Futures Markets*, 35(9), pp. 868–891. doi: 10.1002/fut.21685.

Nwulu, N. I. (2017) 'A decision trees approach to oil price prediction', *IDAP 2017 - International Artificial Intelligence and Data Processing Symposium*, pp. 0–4. doi: 10.1109/IDAP.2017.8090313.

Patalay, S. and Bandlamudi, M. R. (2021) 'Gold Price Prediction Using Machine Learning Model Trees', *International Conference on Changing Business Paradigm, (ICCBP) 2021, MDI Murshidabad*, (October), pp. 1–12.

Ping, P. Y., Miswan, N. H. and Ahmad, M. H. (2013) 'Forecasting malaysian gold using GARCH model', *Applied Mathematical Sciences*, 7(57–60), pp. 2879–2884. doi:

10.12988/ams.2013.13255.

Qian, Y., Ralescu, D. A. and Zhang, B. (2019) 'The analysis of factors affecting global gold price', *Resources Policy*, 64(August), p. 101478. doi: 10.1016/j.resourpol.2019.101478.

Sadorsky, P. (2021) 'Predicting Gold and Silver Price Direction Using Tree-Based Classifiers', *Journal of Risk and Financial Management*, 14(5), p. 198. doi: 10.3390/jrfm14050198.

Shankar, P. S. and Reddy, M. K. (2021) 'Forecasting Gold Prices in India using Time series and Deep Learning Algorithms', *International Journal of Engineering and Advanced Technology*, 10(5), pp. 21–27. doi: 10.35940/ijeat.d2537.0610521.

Singhal, S., Choudhary, S. and Biswal, P. C. (2019) 'Return and volatility linkages among International crude oil price, gold price, exchange rate and stock markets: Evidence from Mexico', *Resources Policy*, 60(September 2018), pp. 255–261. doi: 10.1016/j.resourpol.2019.01.004.

SunshineProfits.com (2022) *Bond Yields*. Available at: https://www.sunshineprofits.com/gold-silver/dictionary/bond-yields-gold/ (Accessed: 25 May 2022).

Torki, L., Samadi, S. and Safrpoor, Z. (2021) 'Analysis of the Effect of Macroeconomic Variables on Fluctuation of Future Gold Market in Iran', 2(1), pp. 251–272. doi: 10.29252/jep.2.1.251.

Uzair, M. and Jamil, N. (2020) 'Effects of Hidden Layers on the Efficiency of Neural networks', *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*, pp. 1–6. doi: 10.1109/INMIC50486.2020.9318195.

Wang, Y., Liu, L. and Wu, C. (2020) 'Forecasting commodity prices out-of-sample: Can technical indicators help?', *International Journal of Forecasting*, 36(2), pp. 666–683. doi: 10.1016/j.ijforecast.2019.08.004.

Wang, Y. and Witten, I. H. (1997) 'Induction of model trees for predicting continuous classes', *Proceedings of the 9th European Conference on Machine Learning Poster Papers*, pp. 128–137. Available at: http://researchcommons.waikato.ac.nz/handle/10289/1183.

Yang, X. (2019) 'The Prediction of Gold Price Using ARIMA Model', *Advances in Social Science, Education and Humanities Research*, 196(Ssphe 2018), pp. 273–276. doi: 10.2991/ssphe-18.2019.66.

Yin, L. and Yang, Q. (2016) 'Predicting the oil prices: Do technical indicators help?', *Energy Economics*, 56, pp. 338–350. doi: 10.1016/j.eneco.2016.03.017.

Zhai, Y., Hsu, A. and Halgamuge, S. K. (2007) 'Combining news and technical indicators in daily stock price trends prediction', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4493 LNCS(PART 3), pp. 1087–1096. doi: 10.1007/978-3-540-72395-0_132.