Comparative Analysis of Machine Learning Algorithms For XAU/USD Prediction: Integrating Economic Indicators And Sentiment Analysis

MSc Research Project
MSc In Data Analytics

Manoj Crasta
Student ID: x23199156

School of Computing
National College of Ireland

Supervisor: Abdul Shahid

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Manoj Crasta
Student ID:	x23199156
Programme:	MSc In Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Abdul Shahid
Submission Due Date:	12/12/2024
Project Title:	Comparative Analysis of Machine Learning Algorithms For XAU/USD Prediction: Integrating Economic Indicators And Sentiment Analysis
Word Count:	7009
Page Count:	20

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

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

Signature:	
Date:	28th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Comparative Analysis of Machine Learning Algorithms For XAU/USD Prediction: Integrating Economic Indicators And Sentiment Analysis

Manoj Crasta
x23199156

Abstract

XAU/USD which is an highly traded forex pair in the financial market, is known for its volatility. Its dependence on diverse factors like economic indicators and sentiment-driven market dynamics, makes it a difficult pair to predict. While existing studies focus on either quantitative data or sentiment data. This study sought to fill this gap by integrating both historical indicators and sentiment scores from news articles to predict XAU/USD hourly rate. The prominent machine learning models were implemented and evaluated using metrics like MAE, MSE and R^2. Random Forest was found to be the most efficient in terms of accuracy and interoperability with Mean Absolute Error of 7.2035. Deep learning models even though they are designed for sequential data were outperformed by ensemble models. While sentiment score contributed to the predictive capability of models, their influence was limited. This research can help the traders and financial analyst to effectively predict the XAU/USD trends by providing a reliable framework. Nevertheless, some additional investigation of sentiment-driven features and real-time analysis tools is necessary to enhance model generalizability and precision.

1 Introduction

The foreign exchange(forex) market which is the most liquid and dynamic market in the global financial system has a daily trading volume exceeding $2 trillion as of April 2022 Chaboud et al. (2023). Forex market which facilitates international trade, investment and speculation acts as a cornerstone of the global economy. Day traders and financial institutions are engaged throughout the day buying, selling and speculating the instruments to earn profit in the forexmarket. Various instruments are traded in the forex market, among these gold, which is traded as XAU/USD, holds a unique position because of its role as a commodity as well as an financial instrument. XAU/USD reflects the price of one ounce of gold in US dollars and gold's properties like corrosion resistance and durability has always upheld its value Andriani et al. (2023). XAU/USD is currently a critical indicator for policymakers and multinational corporations due to its role of safe haven asset and hedge against inflation and economic instability Kilimci (2022). XAU/USD rate prediction is not as simple as it may seem because this currency pair is sensitive to a number of external factors such as geopolitical occurrences, political and other economic activities and even the sentiment of the market. Although it is a complex job, it is considered as essential for financial decisions. As a result of the technological development, the participation in the forex market is rising dramatically. On one side the forex market offers a lot of growth opportunities but on the other side it involves high risk due to its volatility. Hence predicting the instruments like XAU/USD requires diverse data sources from historical data to the market sentiments. This complex nature of XAU/USD has led to many studies and research over the years on financial predictive analytics.

Many studies have been conducted on predicting financial instruments like XAU/USD rate, exploring different methods and diverse datasets focusing on improving prediction accuracy. Some researchers implemented traditional time-series models, such as ARIMA, SARIMA and GARCH and effectively captured the seasonal trends but due to the dynamic nature of financial markets, struggled to learn the non-linear data Pant and Pant (2024);Guha and Bandyopadhyay (2016). This issue was addressed by using machine learning models like SVM, Decision trees and ensemble techniques but temporal dependencies were not caught by these models Sudimanto et al. (2021);Mahajan et al. (2023). Deep learning models were also studied and models like LSTM, GRU performed well for XAU/USD prediction Gong (2024);Birdi et al. (2023). Some previous studies also emphasized on the effects of economic indicators and market sentiment on the instruments. The majority of the studies use either economic indicators or sentiment data, which limits the robustness and comprehensiveness of these models.

The multifaceted nature of affecting factors, such as economic indicators and sentiment data, makes the XAU/USD rate prediction a complex task. Current studies are either focused on quantitative data or sentiment analysis, which fails to capture the complementary nature between these data sources. Moreover, there are very few published works with comparison of different types of machine learning models that uses both economic indicators and sentiment analysis which further can give a better understanding of the effective techniques for forecasting of the XAU/USD rates. These limitations highlight the need of using multiple data sources in the analysis that could include quantitative and qualitative data to build predictive models for XAU/USD rates.

1.1 Research Question

This study addresses the following research question:

How do different machine learning models compare in their ability to predict the XAU/USD exchange rate using a combination of textual sentiment data and various economic indicators?

1.2 Research Objectives

To address the research question, the following specific research objectives were derived:

1. To develop a comprehensive dataset which consists of hourly sentiment scores derived from news articles related to XAU/USD rates, economic indicators and the historical XAU/USD data.

2. To build different machine learning and deep learning models to predict the hourly rates of XAU/USD and evaluating each of those with metrics such as Mean Absolute Error (MAE), Mean Absolute Error (MAE) and R^2 value.

3. To compare different machine learning models based on the performance, robustness and operational efficiency in predicting XAU/USD rates and identifying the best technique.

Hence this study aims to contribute to the field of financial predictive analytics by using integrated data consisting of economic indicators and sentiment analysis to improve robustness and accuracy of XAU/USD forecasting models.

1.3 Proposing Research Model

In this study financial news articles are collected from Perigon API and sentiment analysis is applied to calculate the sentiment scores. Then these hourly sentiment scores are aligned with historical XAU/USD data as well as economic indicators like S&P 500, Crude Oil price, and volatility index to create a comprehensive dataset. Various machine learning and deep learning models including Random forest, XGBoost, LSTM will be trained on this dataset. Best performing model will be identified by evaluating all the models based on the performance metrics. Hence this study will offer valuable insights for traders, investors and policymakers by combining quantitative economic indicators with qualitative sentiment analysis.

1.4 Document Structure

This thesis is structured in a way to provide a logical and clear progression of the research. The first chapter gives an introduction to the research outlining the objectives of the research and motivation. A comprehensive literature review will be presented in chapter 2 which summarizes the traditional methods, , machine learning techniques, and the integration of economic and sentiment data in financial forecasting. In the chapter 3 the methodology which will be followed in the study will be discussed in detail. Data mining techniques followed in data collection, preprocessing, sentiment analysis and model building will be discussed. In the chapter 4 on Results and discussions includes all the interpretations and comparative analysis of the models employed. Chapter 5 concludes the thesis by summarizing the key findings, contributions, and recommendations for future research efforts.

2 Related Work

The related work section examines and discusses various studies on forex market prediction. These studies are organized onto distinct categories. The studies involving traditional forecasting techniques are discussed in the first segment. The subsequent sections discuss the use of machine learning and deep learning techniques in financial market predictions, highlighting their growing relevance in the field. The fourth section explores the studies where economic indicators were included in the financial prediction. It also investigates the influence of sentiment analysis, showing how market sentiments derived from the news impacts the forex trends. In the final section, the gaps in the existing studies will be identified and research niche is outlined. This systematic approach will build a comprehensive foundation on financial predictions while positioning the proposed research to address the current gaps effectively.

2.1 Evolution of Forecasting Techniques

The techniques used for predicting financial markets have been evolved over time from traditional methods to modern machine learning techniques due to the ongoing research in the field. During initial days traders were using fundamental analysis and technical analysis for financial forecasting. Wafi et al. (2015) explored in depth on fundamental analysis methods for predicting stock prices. They focused on models that assess factors like market trends, financial statements and competitive positioning. However, significant drawbacks were identified, which included availability of quality data, dependence highly on market efficiency and rapidly changing market conditions. Also, fundamental analysis proved to be time consuming. Another widely used method for financial predictions is technical analysis which depends on historical price pattern and trading volumes. Neely and Weller (2011) explored the evolution of technical trading rules (TTRs) in forex market and they found that over the time its effectiveness declines due to market dynamics. Their findings also pointed out the need to continuously enhance the forecasting methods. Similarly, Neely (1997) discussed the application of technical analysis in forex market and the demonstrated that although forex traders could earn profits, the ability to do so declines as other traders adopt the same strategies.

To address the challenges of fundamental analysis in short-term variations, along with the declining effectiveness of technical analysis led to the adoption of time-series forecasting methods. Bai (2024)) used the ARIMA(1,1,2) model to predict gold futures prices. Data between 2020 to 2024 was used. The model established a constant R^2 of 0.313 as well as an R^2 of 0.748, and the residual tests did not reveal autocorrelation, indicating white noise. The Ljung-Box statistics (12.286 with p=0.657) also supported validity of the model. The study forecasted gold prices to remain relatively constant in the first half of the year and to rise sharply in the second half of the year and it emphasized the importance of ARIMA models to capture short-term trends. One of the studies by Bunnag (2023) experimented with ARIMA, ARIMA-GARCH and ARIMA-TGARCH models for forecasting daily gold prices. The hybrid model ARIMA(2,1,3)-GARCH(1,1) outperformed other models achieving the lowest MAE of 106.712 and RMSE of 126.788. This demonstrated the effectiveness of ARIMA-GARCH models in forecasting gold prices under volatile market conditions. Guha and Bandyopadhyay (2016) employed the ARIMA model to predict gold prices in India ranged from 2003 to 2014. They implemented six ARIMA model combinations and the best performing model was ARIMA(1,1,1) with R-squared of 0.993 and MAE of 477.33. This study highlighted the limitations of ARIMA in capturing sudden changes caused by economic instability. This research therefore emphasizes use of time-series modeling in prediction of the gold price, while acknowledging its constraints in volatile market conditions. Syrris and Shenai (2021) checked the performance of ARIMA and hybrid ARIMA-GARCH models for gold prices for the period of 2018 to 2019. The ARIMA(4,0,15)-GARCH(1,1) model with normal error distribution (NED) was identified as the most efficient model with the lowest AIC (-7.4459) and the best Theil coefficients. It also explained the combined effect of ARIMA-GARCH where GARCH captured volatility clustering, while ARIMA handled autoregressive processes.

While fundamental and technical analysis showed great results in earlier stages, due to markets dynamic nature its effectiveness declined. Likewise, the time-series models like ARIMA, GARCH and hybrid models failed to address the non-linearity present in the data. This highlights the necessity for employing advanced machine learning models.

2.2 Machine Learning and Deep Learning

This below section explores various studies undertaken in the field of financial prediction using machine learning and deep learning techniques.

Makala and Li (2021) performed a comparison analysis of SVM and ARIMA in predicting the gold prices, with SVM emerging as best one. The study used daily gold price data, from the World Gold Council for 41 years (1979–2019), and imputed missing values by calculating the average of gold price for three consecutive days. The dataset was then split into the training set up to year 2014 and test data from the year 2015 onward. SVM achieved an RMSE of 0.027 and R-squared of 99% where as ARIMA registered RMSE of 36.2. The study Mahajan et al. (2023) employed ensemble techniques for gold price prediction by combining models like decision trees, SVM and gradient boosting. Daily gold price data from 2011 to 2019 was used and compared to individual model's ensemble model performed better. The results showed an MAE of 0.5227, R-squared of 0.8877 and RMSE of 0.7107. This study shed lights on the potential of ensemble models.

A similar study that used ensemble regression models for the prediction of XAU/USD is Kilimci (2022), where data from July 2019 up to July 2020 was used. Technical indicators were also included like moving averages, Bollinger bands etc. The study experimented with the help of different regression models, such as linear regression, decision tree regression, random forest regression, and support vector regression, and some specific ensemble methods voting and stacking regression. The stacking regression model achieved the best results with a MAPE of 2.2036, outperforming other models in predicting gold prices. Sudimanto et al. (2021) compares the performance of SVM, K-Nearest Neighbors, Decision Trees and Ensemble models for the prediction of XAU/USD prices using data from 2019 to 2021. Models were trained on data from 2019-2020 and tested on the 2021 data. It included technical indicators like Moving Average Convergence/Divergence (MACD) for prediction. Coarse Gaussian SVM and 10-NN models showed highest accuracy of 88.9%.

The comparison study N et al. (2021) Linear regression was compared with different types of LSTM (Long Short Term Memory) models such as Vanilla LSTM, Stacked LSTM, and Bidirectional LSTM for global gold prediction from year 1978 to 2018. Study demonstrated that linear regression showed higher errors, with a MAPE of 10.94 and RMSE of 0.153. Vanilla and Stacked LSTM models were the best among all the models that were developed as they yielded the lowest MAPE of 2.649 and RMSE of 0.0292 and 0.0291 respectively. The study highlighted that potential of LSTM in forecasting gold prices, which outperforms linear regression.

Another study S and S (2020) employed LSTM-based deep learning model to predict daily gold prices using data from January 1979 to July 2020. LSTM model performed brilliantly achieving RSME of 7.385 on test data, demonstrating strong alignment between predicted and actual prices. It demonstrated the ability of LSTM over conventional methods such as ARIMA, SVR and CNN for analyzing golds non-linear price fluctuations. Gated Recurrent Unit (GRU) is also evaluated in order to predict gold prices by Sudiatmika et al. (2024). K-fold cross-validation was performed with several combinations of timesteps, batch sizes and epochs. The model with best parameters achieved the lowest MSE of 0.0000324, MAE of 0.012846 with batch size of 64 and 100 epochs. This study showed the potential of GRU for gold prediction.

2.3 Role of Economic Indicators AND Sentiment Analysis

The studies which have used economic indicators and sentiment analysis in predicting financial markets are discussed below.

The study Sami and Nazir (2018) artificial neural networks and linear regression to predict daily gold prices. It used a broader features like 22 market variables like oil prices, S&P 500 index along with daily gold prices from 2005 to 2016. ANN achieved an RMSE as low as 19 outperforming linear regression. This study is notable for incorporating economic indicators, including gold-producing companies' stock values and Russia's interest rates, for forecasting. Another study Cohen and Aiche (2023) explored the application of Random Forest, GBRT and XGBoost to predict gold price fluctuations using data from February 2011 to February 2023. Feature set included global stock indices, commodity futures, VIX volatility and bond yields. Also lagged values up to day 10 was used. The result shows that XGBoost model has the highest accuracy with test MSE of 0.0000186 among other models. The study also pointed out that the three day moving average of gold, lagged stock indices and VIX scores with one day lag significantly affected the prices.

To explore the relationship between gold prices and economic indicators like stock market indices, crude oil prices, exchange rates, inflation and interest rates Manjula and Karthikeyan (2019) performed an regression analysis on gold prices. The study divided the dataset into two distinct periods: January 2000–October 2011 (rising trend) and November 2011–December 2018 (horizontal trend). Out of the three models random forest achieved the best prediction accuracy for entire period with R-Square of 0.9802. The study was also able to establish that the correlation between gold price and its influencing factors was higher during the rising trend period. Rakshitha et al. (2024) utilized machine learning models, including decision trees, regression and ensemble methods for gold prices from December 2016 to December 2023. It also included economic indicators like crude oil prices, metal indices with Adjusted Close price as the dependent variable. Gradient Boosting achieved the lowest RMSE of 0.8433 after feature selection and with R-square of 0.881. It highlighted the influence of cross-validation in enhancing models accuracy.

The effects of financial news sentiment on gold price prediction was discussed by Junjie and Mengoni (2020). The study compared 5-days news sentiment with 1-day news sentiment and 5-days news sentiment had better prediction accuracy of 78.43% and a lower L2 loss at 0.0044. Results from this study showed that there was a negative correlation between positive news sentiment and gold prices to underscore the role of the precious metal as a safe haven or hedge against financial risk. This study demonstrated that using sentiment analysis can enhance the prediction ability of machine learning models. The study Darapaneni et al. (2022) utilized both sentiment analysis and macroeconomic indicators to predict Indian companies stocks and it resulted in accurate prediction. LSTM and Random Forest was used with historical prices of Reliance, HDFC Bank, TCS and SBI companies. Economic indicators like gold prices, Brent Crude, USD-INR exchange rates were included in the dataset along with sentiment analysis of financial news headlines. The LSTM model showed a lowest RMSE of 7.89 for SBI stock with only historical data. While random forest fed with the integrated data and it showed similar RMSEs with the LSTM model. This study therefore establishes the need for the inclusion of financial news sentiment as well as macroeconomic factors in order to improve market forecasting.

2.4 Research Niche

After a comprehensive review of previous studies, a gap was identified in the integration of economic indicators and sentiment analysis for predicting the hourly rate XAU/USD. A Figure 1 in table format summarizes existing studies with remarks.Although there are many studies that focus on either economic indicators or sentiment analysis independently, only a limited number of works that use both qualitative and quantitative data for the prediction of the financial market. In addition, most of the prior research employ daily price data which fails to capture the finer, real-time market fluctuations. Furthermore, there is no prior study which investigates the combined impact of sentiment analysis and economic indicators on XAU/USD despite the significance of this index in the global financial system. This research seeks to fill this gap, by using modern machine learning models that incorporate both economic variables and sentiment analysis providing a comprehensive and robust framework for predicting the XAU/USD rate.

Authors	Target Variable	Features/Variables Used	Models/Techniques Used	Remarks
Mahajan et al. (2023)	Gold Price	Daily gold prices, decision tree, SVM, gradient boosting	Ensemble models (Decision tree, SVM, Gradient Boosting)	Demonstrates the power of ensemble models but lacks macroeconomic factors.
Cohen & Aiche (2023)	Gold Price	Global indices, bonds, commodities, VIX volatility index	Random Forest, GBRT, XGBoost	Incorporates macroeconomic indicators but omits sentiment analysis.
Kilimci (2022)	Gold Price (XAU/USD)	Technical indicators (SMA, RSI, Bollinger bands)	Stacking Regression, Linear Regression, Support Vector Regression	Highlights feature engineering but no use of macroeconomic indicators.
Sudimanto et al. (2021)	FOREX XAU/USD	Moving Average Convergence/Divergence (MACD)	Coarse Tree, Coarse Gaussian SVM, Ensemble Boosted Tree	Focused on technical indicators; lacks macroeconomic and sentiment analysis.
Darapaneni et al. (2022)	Stock Prices	Stock prices, macroeconomic indicators, sentiment data	LSTM, Random Forest	Combines macroeconomic and sentiment data but for stock prices
Vidya & Hari (2023)	Gold Price	Daily gold prices, exponential trend	LSTM	Focuses on LSTM without exploring hybrid models or additional features.
Sudiatmika et al. (2024)	Gold Price	Historical gold prices, normalization	GRU	Focused on GRU models but no use of economic indicators
Junjie & Mengoni	Spot Gold Price	Financial news sentiments, historical gold prices	MLP, VADER Sentiment Analysis	Focused on sentiment data but no broader financial indicators.
Danielsson & Gramer (2021)	EUR/USD Exchange Rate	Sentiment scores, technical indicators	Stacked LSTM, Sentiment Analysis (Longformer, FinBERT)	Combined sentiment and technical indicators but used vast number of predictors

Figure 1: Existing Studies Summary

3 Methodology

This section explains in detail the methodology undertaken in the study. CRISP-DM approach is employed in this study covering all the steps from data collection and data preprocessing and visualization, feature engineering, as well as model selection and model assessment. To make accurate predictions on XAU/USD rates this study leverages both time series economic indicators and sentiment analysis. Hence steps taken to calculate the sentiment scores have also been discussed in this section. This methodology follows a systematic approach first by understand, preparing and analyzing the diverse datasets and then employs different machine learning models to derive actionable insights.

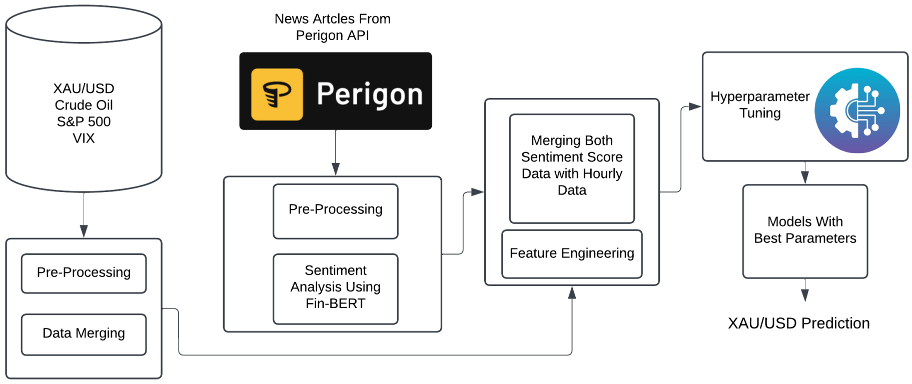

Figure 2: Architecture For XAUUSD Prediction

3.1 Data Collection

The first step in this research involved in collecting the necessary data essential for modelling and analysis. For this study, news articles related to XAU/USD were sourced from Perigon API, which is an HTTP REST API built to fetch news articles and web content. This API delivers responses in JSON format while employing response codes of standard HTTP to provide the status of each request. As Perigon provides data only from July 2020, for this study our datasets are ranged from July 2020 to December 2023. There JSONs containing the news articles were fetched using Postman. All the JSON files were stored systematically in an folder for further processing.

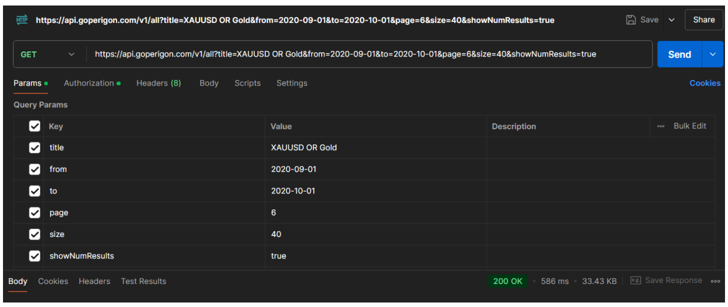

Figure 3: API request in Postman for fetching XAU/USD-related articles

Besides the news articles, hourly historical rate of XAU/USD for same duration was collected from MetaTrader5, which is an recognized trading platform. Along with this hourly data of Crude Oil, S&P500 index and VIX (CBOE Volatility Index) were obtained from the site BacktestMarket.com, which is a reliable source for financial historical data. These historical datasets collected in CSV format and included primary features like Date/Time, Open, Close, Low, High and Volume. These economic indicators were selected on based on their influence on XAU/USD rate.

3.2 Data Pre-processing

In every data mining project preparing the data plays a pivotal role. Similarly in our research various data preprocessing steps undertaken. It involved in handling the news articles for sentiment analysis and then aligning the sentiment scores with the historical financial datasets for robust modelling.

The first step in the data processing phase is extracting and cleaning the news articles which are encoded in JSON files. These files were accessed one by one and two features were extracted with JSON key "pub_date" and "description". The pub_date date has published date and description has article description. The format of pub_date was changed from ISO to standardized date-time format and rounded to the nearest hour for easy calculation of hourly sentiment score in later steps. Furthermore, the duplicates records were removed based on both published date and description. Also the missing or invalid values were replaced with NaN values which were then dropped from the dataset. This cleaned data containing features pub_date and description was then exported from the Pandas dataframe to excel file for further analysis.

In the next step sentiment analysis was performed on the cleaned news data using FinBERT (Financial BERT) model. FinBERT is a pre-trained natural language processing model trained mainly on financial data and designed for financial text analysis. The description of each articles is then fed to the FinBERT pipeline, which generates a sentiment label with a confidence score. FinBERT classifies text into three sentiment categories: positive, negative or neutral.

To make the representation of sentiment simple, a single sentiment score was obtained by consolidating individual scores of positive, negative and neutral. Neutral scores were replaced by 0, positive scores were kept unchanged and negative score were multiplied by -1 to create an unified column named Sentiment Score.

To align with the hourly historical data of XAU/USD the Sentiment Score values are aggregated on hourly basis. The date column pub_date values were rounded to the nearest hour and sentiment scores of all articles published within the same hour are summed. This step provides a summarized score for each hour and this data will be saved in excel for further modelling process.

date	Sentiment Score
2020-07-01 07:00:00	1.569820881
2020-07-02 07:00:00	-0.9695016146
2020-07-03 07:00:00	0.703158319
2020-07-04 07:00:00	-0.3144655228
2020-07-05 07:00:00	-0.9042447209
2020-07-06 05:00:00	0
2020-07-06 07:00:00	0.3261421323
2020-07-07 07:00:00	-0.3172252774
2020-07-07 13:00:00	0.922167182
2020-07-08 07:00:00	2.512453556
2020-07-09 07:00:00	1.206641793

Figure 4: Hourly aggregated sentiment scores

The next steps individual hourly datasets such as XAU/USD, Crude Oil, S&P500 and VIX will be handled. The XAU/USD CSV file has two columns for date and time, which is merged into a single column named TIMESTAMP for easy integration with Sentiment score dataset. Unwanted columns such as OPEN, LOW, HIGH, VOL,SPREAD and TICKVOL were dropped. As the next step datasets XAU/USD and hourly sentiment score dataset are merged using the TIMESTAMP column in the XAU/USD and date column in the Sentiment Score dataset. After merging the dataset there were some missing values in the sentiment score column because no articles were published during those hours. Forward-filling technique was used to handle these missing values. This merged dataset then saved in excel for further process.

Same as XAU/USD the hourly historical data of Crude Oil, VIX and S&P500 were also preprocessed. Date and time column are merged into the single column. All the unwanted columns are dropped retaining only datetime and close columns. To match the date range of XAU/USD, these datasets were filtered to include data between July 2020 and December 2023. There is a column named CLOSE in all the datasets, so this column in each of the datasets is renamed to crude_close, vix_close and sp_close. Next the all three datasets were merged sequentially with the XAU/USD-Sentiment Score integrated dataset using the datetime column as key. This will ensure that all datasets were aligned with each row representing the hourly observation for all variables. After merging missing values are checked and addressed. The final merged dataset provides a unified view of economic indicators and sentiment scores and is shown in the figure below.

CLOSE	Sentiment Score	datetime	crude_close	vix_close	sp_close
1767.92	1.569820881	2020-07-02 05:00:00	40.09	29.33	3124.75
1768.89	1.569820881	2020-07-02 06:00:00	40.15	29.12	3127.25
1767.7	-0.969501615	2020-07-02 07:00:00	40.29	28.25	3146
1766.17	-0.969501615	2020-07-02 08:00:00	40.48	28.27	3153
1767.86	-0.969501615	2020-07-02 09:00:00	39.91	28.95	3128.25

Figure 5: Final Dataset

3.3 Exploratory Data Analysis

To understand more about the selected features and their relationship with the XAU/USD closing price, exploratory data analysis(EDA) was performed. The correlation analysis in Figure 6 showed a negative correlation between the VIX index and XAU/USD, which was contrary to the traditional safe-heaven narrative. This is because during the time between 2020 and 2023, market was recovering from the Covid-19 pandemic effects and also interest rate hikes also might have influenced the relationship. Crude oil showed a weak negative correlation and sentiment score also demonstrated a isolated impact of XAU/USD. S&P 500 demonstrated a slightly positive correlation reflecting a nuanced interplay between gold and equity market trends during the period.

To understand the XAU/USD rate trend and pattern, a time-series graph was plotted Figure 7. It revealed significant fluctuations with sharp increases and decreases. It highlighted the volatility of the gold prices influenced by economic, geopolitical and market factors. There is a long-term upward movement shows the consolidation and correction.

3.4 Feature Engineering

Feature engineering was essential for this study to ensure the data was effectively prepared for predictive modelling. As discussed during the preprocessing all the datasets were formatted into date-time format to easily merge the datasets. Since scaling features

Figure 6: Correlation Heatmap

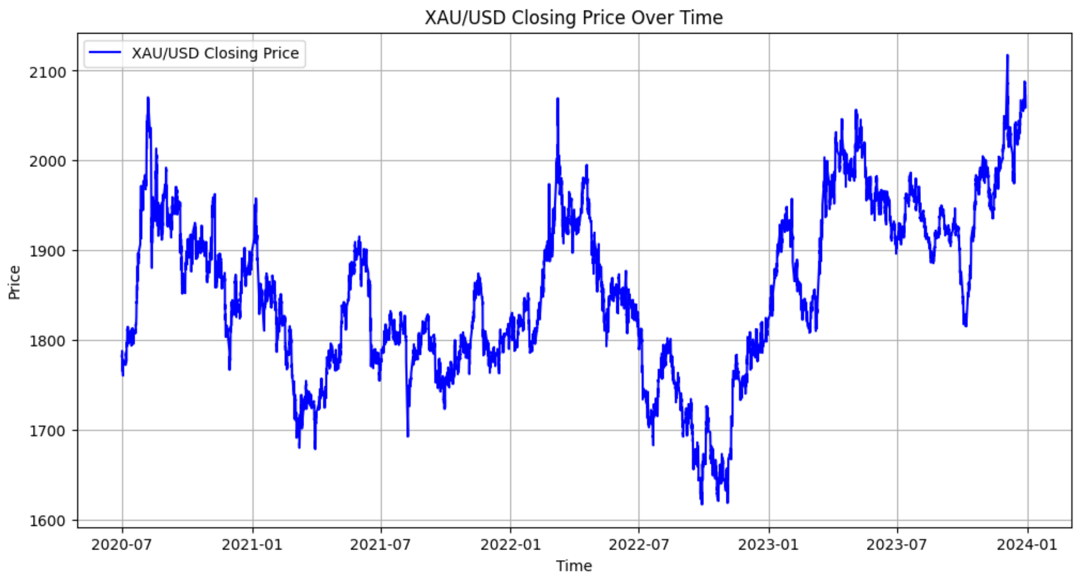

Figure 7: XAU/USD Historical Rates Over Time

facilitates consistency in the data, StandardScaler was applied to transform the data to a standard scale with a mean of zero and a standard deviation of one. The feature CLOSE which represents the hourly closing rate of XAU/USD is selected as target variable, while features such as as sentiment scores, crude oil closing prices, VIX closing values, and S&P 500 closing prices were selected as independent variables. By following these steps the data consistency and suitability for modelling was ensured.

3.5 Modelling

This phase involved in building and training variety of machine learning algorithms to predict the XAU/USD rate. The dataset was split into 80% for training and 20% for testing to evaluate model performance. The algorithms used in this study are Decision Tree, Random Forest, XGBoost, GradientBoost, Supprt Vector Regressor, LSTM and GRU. Each algorithm were selected for its unique strengths ensuring comprehensive modelling and each of these models are described briefly below:

3.5.1 Decision Tree Regressor:

Decision Tree is a basic and yet effective algorithm which divides the data based on a feature thresholds and creates a tree structure to take decisions. Every split thus reduces the variance within the data subsets.

3.5.2 Random Forest Regressor:

Random Forest is a meta algorithm that works on random sample of the data and builds multiple decision trees. The final output will be obtained through an average of the results of the predictions made by each tree. It is also very useful for decreasing overfitting and for increasing the model accuracy.

3.5.3 Support Vector Regressor:

Support Vector Regressor (SVR) was also selected as a kernel-based machine learning model that can detect non-linear structure in financial data. It maps input features to the space of high dimensionality and it determines the correct hyperplane for accurate predictions. Because of its mathematical property and versatility makes it suitable for modeling complicated dependencies between XAU/USD rates.

3.5.4 Gradient Boost Regressor:

Gradient Boosting is a type of ensemble technique which will build weak learners iteratively. Each of the weak learners corrects the errors of its predecessors. It uses gradient descent function to minimize the errors and optimize the performance. It minimizes the errors by using gradient descent function and performs effectively.

3.5.5 XGBoost Regressor:

XGBoost is an optimized version of Gradient Boosting framework which is designed for speed and efficiency. It is well suited to large datasets as it supports parallel computation and it uses regularization techniques to prevent overfitting.

3.5.6 Long Short-Term Memory:

LSTM is a type of RNN which is tailored for sequential data to capture long term dependencies. It achieves through the use of input, forget and output gates to either retain or discard data selectively. LSTMs are known to be superior in time series analysis makes them a suitable for XAU/USD prediction.

3.5.7 Gated Recurrent Unit:

GRU is a simplified LSTM that has less number of parameters yet its performance effectively by capturing the temporal dependencies. GRU simplifies its working by eliminating the complication of multiple gates by as it uses gating mechanisms in order to control the flow of information.

3.5.8 Hyperparameter Tuning:

In our model building phase we have used different methods for hyperparameter tuning as suitable for each type of model. This study employs GridSearch with 5-fold cross-validation for the traditional models such as Random Forest, Decision Tree, Gradient Boosting Regressor, SVR and XGBoost. All the important parameters of each model was explored such as learning rate, maximum depth and number of estimators.

On the other hand for deep learning models LSTM and GRU, we have applied random search to explore hyperparameter space. Here Keras Tuner Library was used for hyperparameters like GRU/LSTM units, activation functions, and learning rate were optimized. Validation loss was used to identify the best hyperparameters. The reason to use distinct approaches here is to ensure the computational efficiency and optimal performance across all model types. The table Table 1 contains all the parameters checked for each models.

4 Evaluation

In this comparative study, standard regression metrics like Mean Squared Error(MSE), Mean Absolute Error (MAE) and R-Squared (R2) were used to evaluate the predictive models to assess their accuracy, error and ability to explain data variability.

1. Mean Squared Error (MSE): It is the square of differences between predicted and actual values. It explains the model's ability to avoid deviations. A lower value indicates high model accuracy and less prediction errors.

2. Mean Absolute Error (MAE): MAE is the average absolute differences between predicted and actual values, , thus providing a clear understanding of prediction accuracy. This is because unlike MSE, it is not skewed by outliers and can therefore be used to determine general trends in error. MAE is an assessment of a model's capability to provide consistent predictions, which means a low MAE is desirable.

3. R-Squared (R^2): R2 represents to what extent the model explains the variance in the target variable. It helps assess the model's ability in capturing the variability in the data.

Table 1: Hyperparameter values for different models.

Model	Parameter	Values
LSTM	lstm_units_1	50, 100, 150, 200
	lstm_units_2	50, 100, 150, 200
	activation	'relu', 'tanh'
	dropout_rate	0.1, 0.2, 0.3, 0.4, 0.5 (step = 0.1)
	learning_rate	1e-4, 1e-3, ..., 1e-2 (log sampling)
	loss	'mse' (Mean Squared Error)
GRU	gru_units_1	50, 100, 150, 200
	gru_units_2	50, 100, 150, 200
	activation	'relu', 'tanh'
	dropout_rate	0.1, 0.2, 0.3, 0.4, 0.5 (step = 0.1)
	learning_rate	1e-4, 1e-3, ..., 1e-2 (log sampling)
	loss	'mse' (Mean Squared Error)
Gradient Boosting	n_estimators	50, 100, 150
	learning_rate	0.01, 0.1, 0.5, 1.0
	max_depth	3, 5, 7
Random Forest	n_estimators	50, 100, 150
	max_depth	None, 10, 20, 30
	min_samples_split	2, 5, 10
	min_samples_leaf	1, 2, 4
Decision Tree	max_depth	3, 5, 10, 20
	min_samples_split	2, 5, 10
	min_samples_leaf	1, 2, 4
XGBoost	n_estimators	50, 100, 200
	max_depth	4, 6, 8
	learning_rate	0.01, 0.1, 0.2
	subsample	0.7, 0.8, 1.0
	colsample_bytree	0.7, 0.8, 1.0
SVR	C	0.1, 1, 10, 100
	kernel	'linear', 'rbf'
	gamma	'scale', 'auto', 0.1, 1, 10

4.1 Experiment 1: Decision Tree Regressor

The Decision Tree Regressor model with best parameters performed reasonably with MSE of 354.7371, MAE of 7.9344 and R^2 of 0.9587. Further feature importance analysis showed that crude oil prices and S&P 500 closing prices were the most significant predictors. The model effectively captured of the significant trends of the data but had some discrepancies, pointing to the fact that even deeper or ensemble based kinds of models may help more due to the data complexity.

4.2 Experiment 2: Random Forest Regressor

Random Forest Regressor performed better than any other model with MSE of 186.9992, an MAE of 7.2035, and an R^2 score of 0.9782. Due to the ensemble capabilities model

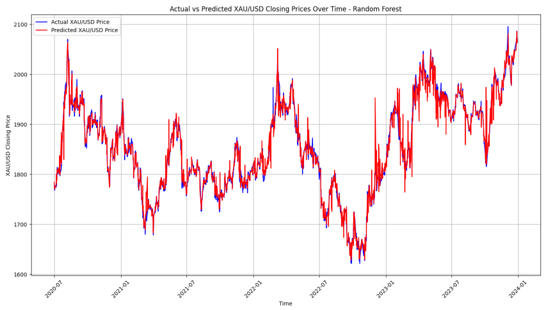

Figure 8: Actual vs Predicted XAU/USD Closing Prices Using Random Forest Model Over Time

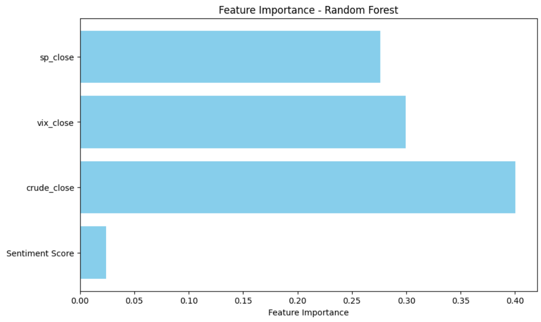

Figure 9: Feature importance plot of the model

improved its ability to generalize and reduced overfitting. Because of its high accuracy and less sensitivity to noise, this model would be suitable for financial forecasting.

The feature importance plot in Figure 9 shows that the features 'crude_close' and 'sp_close' are the most significant predictors, while the 'sentiment score' has a relatively smaller influence on the prediction. Also time-series plot in Figure 8 which compares the actual and the predicted XAUUSD rates shows the capability of the random forest to closely follow the trends and fluctuations.

4.3 Experiment 3: Gradient Boosting Regressor

Gradient Boosting resulted in MSE of 311.2950, MAE of 10.5101 and R^2 of 0.9637. This model was able to progressively rectify its mistakes with the help of the boosting mechanism and therefore performed well. Thus, somewhat higher error metrics compared to Random Forest indicate that some dynamics can be missed due to the choice of hyperparameters or interactions with features.

4.4 Experiment 4: XGBoost Regressor

The XGBoost model achieved a R^2 score of 0.9760 with an MSE of 205.9491 and MAE of 8.5170. The feature importance plot showed that crude oil prices and VIX close prices are the key predictors. This is likely due to the ability of XGBoost to handle non-linearity relationships and interactions which has been suggested in existing studies Cohen and Aiche (2023) where XGBoost models have outperformed other models when dealing with similar financial datasets.

4.5 Experiment 5: Support Vector Regressor

The performance of SVR was moderate in this study, achieving an MSE of 452.8473, an MAE of 11.4575, and an R^2 score of 0.9472. However, SVR, which is known for its performance in non-linear data, struggled with scale of the dataset, which may be due to its to its sensitivity to parameter tuning.

4.6 Experiment 7: Long Short-Term Memory (LSTM)

The LSTM model, which is designed for sequential data, achieved an MSE of 587.2744, an MAE of 17.2771, and an R^2 score of 0.9316. Even though it has temporal capabilities, its performance lagged behind the ensemble models, which implies that feature-level interactions may overpowered the temporal dependencies in this dataset. The Figure 11 shows the training and validation loss curves for LSTM model over the epochs. We can see that the steep decline in both curves reflects the capability of LSTM model to train effectively on our dataset. Also from the plot we can see that LSTM exibits strong generalization and minimized overfitting.

Figure 10: Training and Validation Loss Over Epochs for GRU

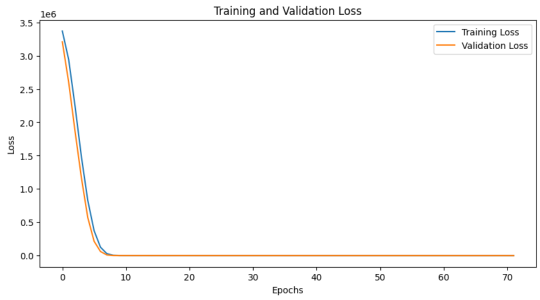

Figure 11: Training and Validation Loss Over Epochs for LSTM

4.7 Experiment 7: Gated Recurrent Unit (GRU)

GRU was the lowest performed model in this study. Even though it was a computationally efficient alternative to LSTM, it achieved MSE of 697.5075, an MAE of 18.9500, and an R^2 score of 0.9187. Despite doing the hyperparameter tuning using RandomSearch, its rapid convergence in the plot Figure 11 suggests the chances of overfitting.

However, these metrics fall short compared to other models like Random Forest, suggesting that GRU's performance may have been affected by dataset constraints or limitations in the hyperparameter tuning process. The model's rapid convergence also suggests potential overfitting or limited learning capacity, emphasizing the need for further experimentation with larger, more diverse datasets and alternative tuning methods to enhance its predictive accuracy.

4.8 Best Performing Model

Out of all the models Random Fores Regressor emerged as the best performing model in terms of both metrics and simplicity. It achieved the lowest MSE and highest R-squared score while maintaining interpretability. The second ranked model was XGBoost because of its strong performance metrics and its capability to handle non-linearity in the data. These results are consistent with previous studies, where ensemble models have shown their capability in similar financial forecasting tasks. This evaluation confirms the use of ensemble models in capturing complex patterns and interactions in the financial datasets. While deep learning models like LSTM and GRU are well suited for sequential data tasks they were outperformed in this study by ensemble methods probably because of the kind of features employed. The table Table 2 shows the best parameters for each model and their metrics.

Table 2: Best Parameters of Models And Their Metrics

Model	**Best Parameters**	**R^2 Score**	**MSE**	**MAE**
LSTM	lstm_units_1: 200, lstm_units_2: 150, activation: relu, dropout_rate: 0.3, learning_rate: 0.00086	0.9316	587.27	17.28
GRU	gru_units_1: 150, gru_units_2: 200, activation: tanh, dropout_rate: 0.4, learning_rate: 0.00922	0.9187	697.51	18.95
Gradient Boosting	n_estimators: 150, learning_rate: 0.5, max_depth: 7	0.9637	311.3	10.51
Random Forest	n_estimators: 150, max_depth: 30, min_samples_split: 2, min_samples_leaf: 1	0.9782	187	7.2
Decision Tree	max_depth: 20, min_samples_split: 2, min_samples_leaf: 1	0.9587	354.74	7.93
XGBoost	colsample_bytree: 1.0, learning_rate: 0.2, max_depth: 8, n_estimators: 200, subsample: 0.7	0.976	205.95	8.52
SVR	C: 100, gamma: 10, kernel: rbf	0.9472	452.85	11.46

4.9 Discussion

In this study I have implemented popular machine learning models for the prediction of XAU/USD hourly rate using economic indicators and sentiment analysis. The results showed that most models demonstrated significant predictive capabilities but limitations and ways to enhance the results became apparent during the analysis.

After tuning all the models with the best parameters, Random Forest emerged as the most effective model in this study. It showed an balance between metrics and interpretability, proving it to be a robust choice for XAU/USD rate prediction. LSTM was able to capture temporal dependencies well, but due to high computational requirement and extensive tuning makes it unsuitable for real-world, time-sensitive applications. GRU showed overfitting and failed to exhibit generalization. Other ensemble models like XGBoost and Gradient Boosting also showed their capabilities in handling non-linear conditions making them ideal for financial problems. While these models performed well in predicting XAUUSD, but the feature importance plot Figure 9 of all models revealed that impact of sentiment score feature was limited. One of the reasons can be that, in

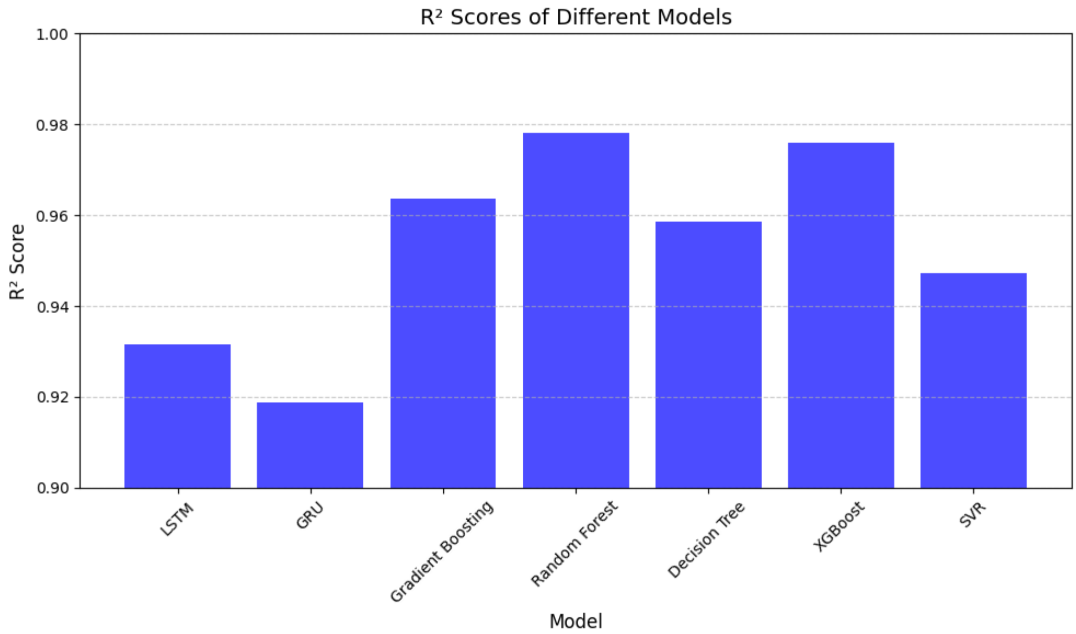

Figure 12: R^2 Scores of All Models

this project Perigon API was used for the news articles collection and this API only had the access to news articles from July 2020. Another reason can be sentiment scores inability to accurately reflect the markets's reaction to news headlines. Because not all news will have direct or proportional impact on financial decisions. The major limitation of this study is the dataset size, due to the dependency on Perigon API as discussed earlier. Although the models performed well, having a larger and more diverse dataset would increase generalization and model reliability. Other features based on sentiments like polarity scores and keywords-based sentiments were not explored due to time constraint. Perhaps, the use of domain specific insights and high frequency financial data could better capture relationships between sentiment and gold prices.

The results of this study aligns with the previous studies that ensemble models have always been superior in compared to other standalone models or deep learning models in the field of financial prediction. This study has contributed valuable insights to the financial field by demonstrating how machine learning models can be employed to predict XAU/USD rates integrating sentiment analysis and economic indicators. Although the study's exhibited limitations due to the limited sentiment data and small data size, it revealed strong predictive capabilities of various machine learning models, especially ensemble models.

5 Conclusion and Future Work

In this study, the primary goal was to perform a comparison analysis of predictive capabilities of various popular machine learning models in predicting XAUUSD rate integrating economic indicators and sentiment from the news articles. To meet these objectives, a comprehensive dataset was developed. It included historical hourly rates of XAUUSD, hourly sentiment scores derived from the news articles and economic indicators namely crude oil, VIX index and S&P 500. Different machine learning models were implemented and compared based on the evaluation metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

Out of all the models implemented Random Forest Regressor emerged as the effective model due to its interpretability and computational efficiency. XGBoost and Gradient Boost Regressor also performed similarly revealing that ensemble models perform best for the similar financial prediction applications.

Although the performance of models in predicting the XAU/USD rate was great, it was observed that sentiment scores had relatively limited influence on the predictions. Future work can be carried out with a broader dataset like including more economic indicators related to XAU/USD and larger number of articles of wider timeframe. Besides these other sentiment features like polarity scores or keyword-specific sentiment score techniques can be explored.

References

Andriani, W., Gunawan and Prayoga, A. (2023). Prediksi nilai emas menggunakan algoritma regresi linear, *Jurnal Ilmiah Informatika Komputer* **28**: 27–35.

Bai, Y. (2024). Research on gold price prediction based on arima model, *Theoretical and Natural Science* **52**: 41–48.

Birdi, N., Sarangi, P. K., Singla, T., Sahoo, A. K. and Subudhi, G. B. (2023). Forex trend analysis using rnn and lstm models: Predictive analytics through neural networks, *2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)*, pp. 206–210.

Bunnag, T. (2023). The importance of gold and forecasting of the gold price using arima, arima-garch, and arima-tgarch model.

Chaboud, A., Rime, D. and Sushko, V. (2023). The foreign exchange market, *Technical Report BIS Working Papers No. 1094*, Monetary and Economic Department. JEL classification: F31, G15.
URL: *https://www.bis.org/publ/work1094.html*

Cohen, G. and Aiche, A. (2023). Forecasting gold price using machine learning methodologies, *Chaos, Solitons Fractals* **175**: 114079.
URL: *https://www.sciencedirect.com/science/article/pii/S0960077923009803*

Darapaneni, N., Paduri, A., Sharma, H., Manjrekar, M., Hindlekar, N., Bhagat, P., Aiyer, U. and Agarwal, Y. (2022). Stock price prediction using sentiment analysis and deep learning for indian markets.

Gong, W. (2024). Research on gold price forecasting based on lstm and linear regression, *HS Web of Conferences* **181**: 02005. ICDEBA 2023.

Guha, B. and Bandyopadhyay, G. (2016). Gold price forecasting using arima model, *Journal of Advanced Management* **4**(2): 117–121.

Junjie, Z. and Mengoni, P. (2020). Spot gold price prediction using financial news sentiment analysis, *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pp. 758–763.

Kilimci, Z. (2022). Ensemble regression-based gold price (xau/usd) prediction, *Journal of Emerging Market Finance* **2**: 7.

Mahajan, R., Patil, P., Chikmurge, D. and Barve, S. (2023). Forecasting gold price using ensemble based machine learning approach.

Makala, D. and Li, Z. (2021). Prediction of gold price with arima and svm, *Journal of Physics: Conference Series* **1767**: 012022.

Manjula, K. A. and Karthikeyan, P. (2019). Gold price prediction using ensemble based machine learning techniques, *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1360–1364.

N, D., Prajapati, P. R., M, R., Ramesh, R., Kodipalli, A. and Joy Martis, R. (2021). Prediction of gold price using deep learning, *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 1–5.

Neely, C. (1997). Technical analysis in the foreign exchange market: A layman's guide, *Review* **79**: 23–38.

Neely, C. and Weller, P. (2011). Technical analysis in the foreign exchange market, *Technical Report Working Paper 2011-001B*, Federal Reserve Bank of St. Louis Research Division. Accessed: 1 August 2024.

Pant, K. and Pant, A. (2024). Exchange rate prediction using time series approach, *Title of the Journal or Conference Proceedings* . Accessed: 2 August 2024.

Rakshitha, Y., Divya Sri, K., Swetha, B., Alajangi, S. and Kavitha, D. (2024). Data mining strategies for gold price prediction using multi-factorial influences, *International Journal for Multidisciplinary Research (IJFMR)* **6**: 1. M.TECH. Integrated CSE with specialization in Business Analytics, VIT Chennai.
URL: *https://www.ijfmr.com*

S, V. G. and S, H. V. (2020). Gold price prediction and modelling using deep learning techniques, *2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS)*, pp. 28–31.

Sami, I. and Nazir, K. (2018). Predicting future gold rates using machine learning approach, *International Journal of Advanced Computer Science and Applications* **8**: 92–99.

Sudiatmika, I., Putra, I. and Artana, W. (2024). The implementation of gated recurrent unit (gru) for gold price prediction using yahoo finance data: A case study and analysis, *Brilliance: Research of Artificial Intelligence* **4**: 176–184.

Sudimanto, Y., Heryadi, L. and Wibowo, A. (2021). Foreign exchange prediction using machine learning approach: A pilot study, *2021 4th International Conference on Information and Communications Technology (ICOIACT)*, IEEE, pp. 239–242.

Syrris, I. and Shenai, V. (2021). Forecasting gold prices with arima and garch models.

Wafi, A. S., Hassan, H. and Mabrouk, A. (2015). Fundamental analysis models in financial markets – review study, *Procedia Economics and Finance* **30**: 939–947. Available at: https://doi.org/10.1016/S2212-5671(15)01344-1 (Accessed: 1 August 2024).