USING STATISTICAL MODEL AND MACHINE LEARNING FOR GOLD PRICE PREDICTION

Doan Hoai Viet   
*STAT3013.P12.CTTT*  
*University of Information Technology*  
22521662@gm.uit.edu.vn Ngo Thanh Trung  
*STAT3013.P12.CTTT*  
*University of Information Technology*  
@gm.uit.edu.vn

lHuynh Quoc Minh Thong   
*STAT3013.P12.CTTT  
University of Information Technology*  
@gm.uit.edu.vn

*Abstract*— This report presents a comparative analysis of three models, including Linear Regression, Gated Recurrent Unit (GRU), Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) for predicting gold prices. The models are evaluated using MAE, MAPE, and RMSE metrics on historical gold price indicators data. The model with the lowest MAE, MAPE, and RMSE is recommended for gold price forecasting, contributing to improved understanding and accurate predictions in the gold market.

Keywords— gold price, forecasting, GRU, ARIMA, LSTM, gc=f, GLD, NEM…

# INTRODUCTION

The prices of gold trading symbols such as GC=F (Gold Futures), GLD (SPDR Gold Shares ETF), and NEM (Newmont Corporation) are significantly influenced by supply and demand dynamics, shaped by factors such as economic conditions, geopolitical uncertainties, inflation rates, and consumer preferences. Historically, gold and related assets have been viewed as safe-haven investments, particularly during periods of rising inflation, when investors seek to protect their wealth, leading to increased demand and higher prices. Conversely, during times of low inflation, demand for these assets may decline, causing prices to drop. Understanding these relationships is essential for identifying market trends and making informed investment decisions. This research focuses on evaluating the effectiveness of various statistical and machine learning models in predicting the future prices of GC=F, GLD, and NEM.

# RELATED WORK

Numerous studies have explored methods for predicting gold prices. For instance, Zhi Li [1] demonstrated the effectiveness of the random forest algorithm, incorporating factors such as the DJIA and S&P500. Shruti Garg [2] employed a Bayesian neural network, achieving a mean percentage error of just 1%. While primarily focused on temperature forecasting, Garima Jan [3] provided a comparison between ARIMA and Exponential Smoothing models. Alex Sherstinsky [4] offered an in-depth tutorial on LSTM networks. In Vietnam, Ho Thanh Tri et al. [5] applied the ARIMA(5,1,5) model to forecast gold prices, while Dan Liu and Xiaohui Yang [6] identified ARIMA(3,1,2) as the optimal model for their predictive analysis.

# ARTIFICIAL INTELLIGENCE MODELS

To forecast the prices of silver, gold, and platinum, our methodology integrates a combination of machine learning and statistical techniques. These includeGated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), Autoregressive Integrated Moving Average (ARIMA),

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## GRU

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

To solve the vanishing gradient problem of a standard RNN, GRU uses, Update gate and Reset gate. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or removing information which is irrelevant to the prediction.

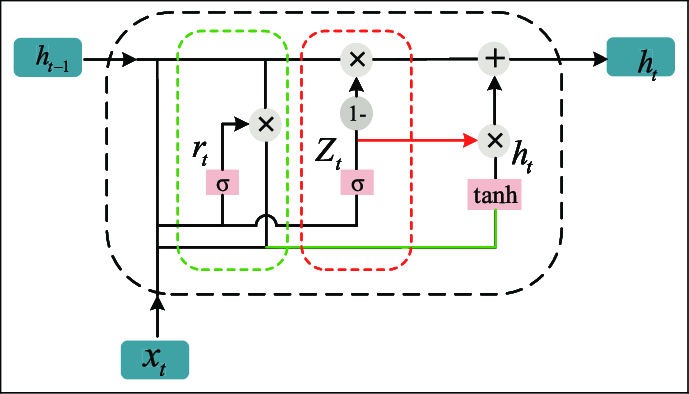


Figure . Gru model

The update gate helps the model to determine how much of the past information (from previous time steps) needs to be passed along to the future and avoid the vanishing gradient problem. The formula using for Update gate:

When is plugged into the network unit, it is multiplied by its own weight ( ). The same goes for ℎ which holds the information for the previous t - 1 units and is multiplied by its own weight ( ). Both results are added together, and a sigmoid activation function is applied to squash the result between 0 and 1.

Reset gate is used from the model to decide how much of the past information to forget. To calculate it, we use:

This formula is the same as the one for the update gate. We plug in ℎ — blue line and — purple line, multiply them with their corresponding weights, sum the results and apply the sigmoid function. The difference comes in the weights and the gate’s usage.

To determine what to remove from the previous time steps, we will calculate current memory content by using reset gate to store relevant information from the past.

To process input and ℎ , we multiply them by weights and respectively. The reset gate is used to determine what information to remove from previous time steps by calculating the Hadamard product with ( ℎ ). The results of steps 1 and 2 are summed up and passed through the tanh activation function.

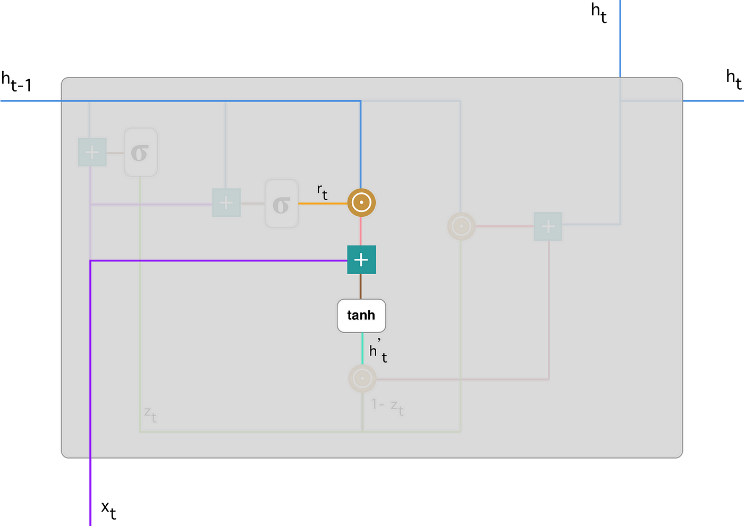


Figure . Current memory content

Finally, the network needs to calculate ℎ ℎ′ — vector which holds information for the current unit and passes it down to the network. In order to do that the update gate is needed. It determines what to collect from the current memory content — and what from the previous steps — ht-1. That is done as follows:

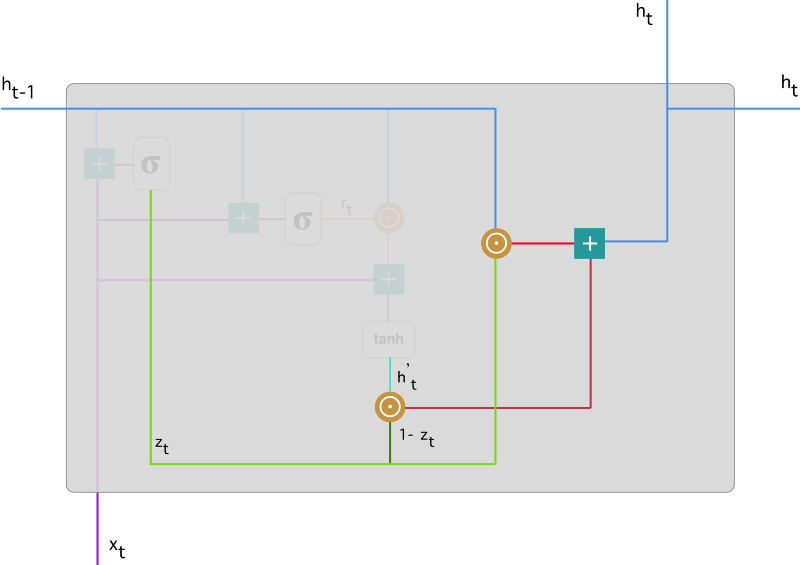


Figure . Final memory content

The model can learn to set the vector close to 1 and keep a majority of the previous information by apply element-wise multiplication to the update gate zt and ℎt-1. Since zt close to 1 at this time step, (1 − zt)will be close to 0 and ignore big portion of the current content, which is irrelevant for our prediction by apply element-wise multiplication to (1 – zt ) and . Final, we sum the results.[23]

## ARIMA

ARIMA is a statistical model that uses time series data to understand the dataset or predict future trends. Autoregressive models assume that the future will resemble the past and predict future prices based on past performance, but it can be inaccurate during certain market conditions. ARIMA analyzes the strength of one dependent variable compared to other changing variables and aims to predict future market moves by examining differences between values in the series.[12]

* The MA component predicts the value by accounting for the past errors or random "shocks" in previous observations. It aims to smooth out the time series by combining the random variations that were not explained.
* Auto regression(AR): The process of finding the relationship between the current data and p past data points (referred to as lags).

𝑦𝑡 : is the value of a dependent variable at time t

𝑎0 : is a constant term

𝑦𝑡−1, 𝑦𝑡−2, 𝑦𝑡−𝑝 : the dependent variables in the past

𝑎1,𝑎2*,* 𝑎𝑝 : are the coefficients of the lagged values of the dependent variables respectively

𝜀𝑡 : is the error term at time t

Do

* Moving average(MA):

The MA component predicts the value by accounting for the past errors or random "shocks" in previous observations. It aims to smooth out the time series by combining the random variations that were not explained.

Where,

𝑦𝑡: is the value of a dependent variable at time t

𝛽0: is a constant term

𝜀𝑡−1, 𝜀𝑡−2, 𝜀𝑡−𝑞: the error terms in the past

𝛽1,𝛽2*,* 𝛽𝑞: are the coefficients of the lagged values of the error term respectively

𝜇𝑡: is the current error term or residual at time t

* ARIMA:

ARIMA(p,d ,q)= AR(p)+ I(d)+ MA(q)[13]

Where,

p, d, q: are non-negative integers representing the order of autoregression, differencing, and moving average respectively.

AR(p): Autoregression of order p refers to the regression of the variable on its own lagged values up to p lags. The AR(p) component captures the linear dependence between the current value and its past values.

I(d): Integration of order d refers to differing the time series d times until it becomes stationary. The I(d) component removes the trend from the time series.

MA(q): Moving average of order q refers to the regression of the variable on the past errors up to q lags. The MA(q) component captures the linear dependence between the current value and its past errors.

## LSTM

LSTM (Long Short-Term Memory) is a deep learning, sequential neural network that overcomes the vanishing gradient problem faced by RNN. It allows information to persist and works similarly to an RNN cell.

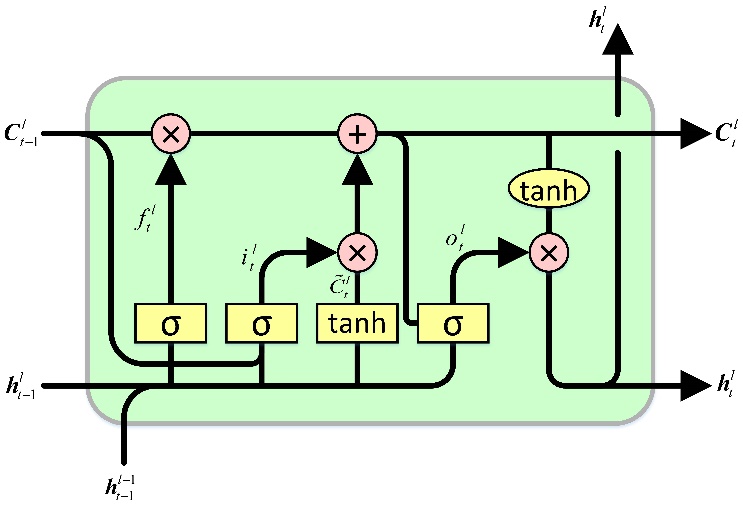


Figure . LSTM Architecture

An LSTM has a hidden state for short-term memory and a cell state for long-term memory. The cell state remembers important information over time while filtering out irrelevant data, allowing LSTMs to make accurate predictions.

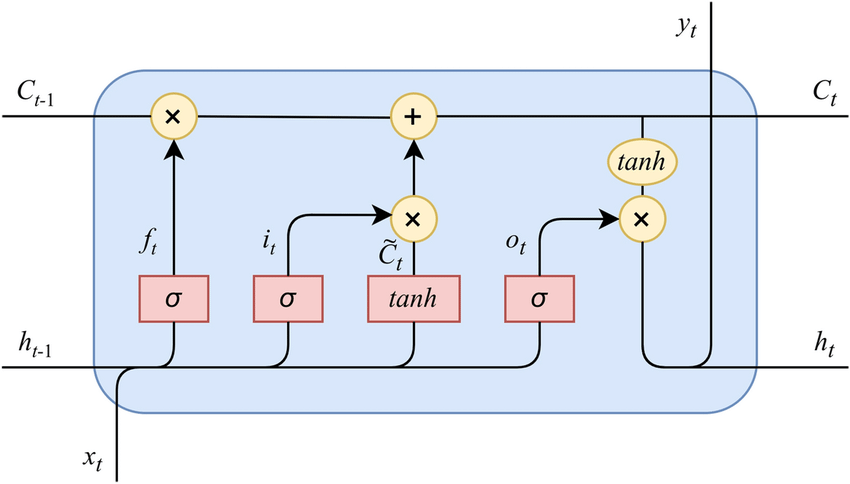


Figure . Long and Short Term Memory

LSTM has three gates that control the flow of information in and out of the memory cell: Forget, Input, and Output.

First, the Forget gate chooses whether the information coming from the previous timestamp is to be remembered or is irrelevant and can be forgotten.

1

𝜎 = 1 + 𝑒−𝑥

𝑓𝑡 = 𝜎(𝑊ℎℎ(𝑓) ℎ𝑡−1 + 𝑊ℎ𝑥(𝑓) 𝑥𝑡)

It determines how much information from the previous cell state 𝐶𝑡−1 should be retained for the current time step. It does this by taking a weighted sum of the previous hidden state ℎ𝑡−1 and current input 𝑥𝑡, with weights given by matrices 𝑊ℎℎ(𝑓)

and 𝑊ℎ𝑥(𝑓), respectively.

The result is passed through a sigmoid activation function σ, which squashes the values between 0 and 1. A value of 0 means that all information from the previous cell state will be forgotten, while a value of 1 means that all information will be retained.

Next, the Input gate tries to learn new information from the input to this cell and determines how much to take from the intermediate cell state.

𝑖𝑡 = 𝜎(𝑊ℎℎ(𝑖) ℎ𝑡−1 + 𝑊ℎ𝑥(𝑖) 𝑥𝑡)

In contract, the Input gate identifies how much information from the current input 𝑥𝑡 and previous hidden state ℎ𝑡−1 should be used to update the cell state 𝐶𝑡, with weights given by matrices 𝑊ℎℎ(𝑖) and 𝑊ℎ𝑥(𝑖), respectively.

Again, we have applied the sigmoid function over it. As a result, the value of 𝑖𝑡 will be between 0 and 1.

𝐶̃𝑡 = 𝑡𝑎𝑛ℎ(𝑊ℎℎ(𝑐) ℎ𝑡−1 + 𝑊ℎ𝑥(𝑐) 𝑥𝑡)

LSTM may not take everything from the intermediate cell state, the 𝐶̃𝑡 is there to distinguish between different weight matrices. Due to the 𝑡𝑎𝑛ℎ function, the value of new information will be between -1 and 1. If the value of 𝐶̃𝑡 is negative, the information is subtracted from the cell state, and if the value is positive, the information is added to the cell state at the current timestamp. However, the 𝐶̃𝑡 won’t be added directly to the cell state.

Here comes the updated equation :

𝐶̃𝑡 = (𝑓𝑡 𝐶𝑡−1 + 𝑖𝑡 𝐶̃𝑡)

The Output gate passes the updated information from the current timestamp to the next timestamp. Together, they function as a layer of neurons with hidden layers and a current state.

𝑜𝑡 = 𝜎(𝑊ℎℎ(𝑜) ℎ𝑡−1 + 𝑊ℎ𝑥(𝑜) 𝑥𝑡)

Its value will also lie between 0 and 1 because of this sigmoid function. Now to calculate the current hidden state, we will use 𝑜𝑡 and tanh of the updated cell state. [21]

ℎ𝑡 = 𝑜𝑡 ∗ tanh(𝐶𝑡)

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.
8. K. Eves and J. Valasek, “Adaptive control for singularly perturbed systems examples,” Code Ocean, Aug. 2023. [Online]. Available: <https://codeocean.com/capsule/4989235/tree>
9. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2013, arXiv:1312.6114. [Online]. Available: <https://arxiv.org/abs/1312.6114>
10. S. Liu, “Wi-Fi Energy Detection Testbed (12MTC),” 2023, gitHub repository. [Online]. Available: https://github.com/liustone99/Wi-Fi-Energy-Detection-Testbed-12MTC
11. “Treatment episode data set: discharges (TEDS-D): concatenated, 2006 to 2009.” U.S. Department of Health and Human Services, Substance Abuse and Mental Health Services Administration, Office of Applied Studies, August, 2013, DOI:10.3886/ICPSR30122.v2

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the confer**