SENTIMENT-DRIVEN GOLD PRICE PREDICTION USING NEWS HEADLINES AND SOCIAL MEDIA

HAZIM FITRI BIN AHMAD FAUDZI

P152419

UNIVERSITI KEBANGSAAN MALAYSIA

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
RF	Random Forest
UKM	Universiti Kebangsaan Malaysia
VADER	Valence Aware Dictionary for sEntiment Reasoning
XGBoost	Extreme Gradient Boosting

CHAPTER I

INTRODUCTION

1.1   RESEARCH BACKGROUND

Gold has long been regarded as a critical financial asset and safe investment, with its price influenced by economic factors and market sentiment. In today's digital age, vast amounts of textual data from news media and social networks reflect investor mood and expectations. Harnessing this sentiment information for predictive modelling is an emerging trend in financial analytics. Sentiment analysis refers to using computer in determining whether textual content conveys a positive, negative, or neutral opinion. News headlines can capture market-moving information such as policy changes, crises, economic indicators, while social media discussions such as tweets or forum posts often reveal investor optimism or fear in real time. Integrating these unstructured data sources with traditional time-series data could potentially improve the accuracy of gold price forecasts.

Recent studies have demonstrated the value of sentiment data in financial prediction. For instance, Bollen et al. (2011) showed that public mood derived from Twitter significantly improved the prediction of stock market movements, achieving about 87.6% accuracy in predicting daily market direction when mood indicators were included. This suggests that collective investor sentiment can serve as a leading indicator for price changes. In the context of commodities like gold, which is often seen as an indicator of economic confidence and uncertainty, sentiment expressed in financial news and social media might likewise contain predictive signal. However, much of the prior research on sentiment-driven forecasting has focused on equities or cryptocurrencies, and relatively few studies have focused on gold prices. Gold's unique role as a hedge against uncertainty means that news about geopolitical events or

economic policy, as well as chatter among investors, may strongly sway its price. This research proposal thus position itself at the intersection of natural language processing, sentiment analysis, and financial time-series modelling, aiming to fill a gap in existing literature by focusing on gold price prediction enhanced with sentiment features.

1.2   PROBLEM STATEMENT

Gold price forecasting remains a challenging task due to volatility and the multitude of influencing factors. Traditional forecasting models often rely purely on quantitative time-series data such as historical prices and macroeconomic indicators. It may ignore the rich information embedded in textual news and social sentiment. This oversight can lead to suboptimal predictions, as market prices are not driven by fundamentals alone but also by investor psychology and reactions to news. The specific gap that this study addresses is the lack of an integrated approach that combines news headline sentiment and social media sentiment with machine learning for gold price prediction.

While prior works have examined sentiment in financial contexts, few have focused on gold or have done so using advanced NLP techniques tailored to finance. For example, Zhou and Mengoni (2020) utilized financial news sentiment for spot gold price prediction, but their work did not incorporate social media data. On the other hand, studies of social media sentiment on markets have typically centred on stock indices or cryptocurrencies. There is a need to investigate whether combining multiple sentiment sources, news and social media can improve predictive performance for gold specifically. Moreover, it is unclear which modelling approach is most suitable to combine sentiment information with price data.

The problem statement for this research can thus be summarized as follows. Firstly, can the inclusion of sentiment indicators derived from news headlines and social media posts significantly enhance the accuracy of gold price prediction models, and what analytical approach yields the best performance in leveraging these sentiment signals? In addressing this problem, the study will also tackle related questions, such as determining the comparative effectiveness of lexicon-based vs. transformer-based sentiment analysis in the financial domain and identifying which type of sentiment between news, social, or a combination has greater predictive value for short-term gold

price movements. The ultimate aim is to bridge the knowledge gap by developing a sentiment-driven forecasting framework for gold that outperforms traditional models lacking such unstructured data inputs.

1.3   RESEARCH OBJECTIVES

Given the above problem, this research aims to achieve a general objective supported by several specific objectives.

1.3.1   General Objective

To develop and evaluate predictive modelling framework for gold price that integrates sentiment analysis of news headlines and social media content, in order to improve forecast accuracy and provide insights into the impact of market sentiment on gold price movements.

1.3.2   Specific Objective

1. To perform sentiment analysis on gold-related news headlines and Twitter posts.

2. To investigate the correlation between sentiment scores and gold price fluctuations.

3. To compare the performance of various machine learning models using sentiment data as features.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION

2.1.1 Sentiment Analysis in Financial Markets

There is an abundant source of literature exploring the relationship between public sentiment and market behaviour. Investor psychology theories state that investor emotions and expectations can drive asset prices away from fundamental values in the short term. One evidence to supports this is Bollen et al. (2011) famously found that certain collective mood states extracted from Twitter feeds had a significant correlation with, and even predictive power for, stock index movements. Similarly, other studies have reported that incorporating sentiment indicators can improve forecasting models for equities and cryptocurrencies. These findings motivate our application of sentiment analysis to gold, hypothesizing that gold will particularly benefit from sentiment-aware models as gold is a commodity that is sensitive to crises and investor fear. Indeed, investor sentiment has been identified as a factor influencing gold price, especially during periods of uncertainty. An example of periods of uncertainty is political turmoil which often increases positive sentiment towards gold as a safe investment.

2.1.2 News Headlines vs. Social Media

Sentiment can be obtained from two main sources which are official news and casual social media. News articles tend to provide factual reporting and expert commentary on economic events, whose tone can influence investor outlook. Prior research on commodities has started to utilize news sentiments. For example, Hajek and Novotny (2022) presented a fuzzy rule-based model combining financial variables with news sentiment, and found that sentiment features significantly improved the accuracy of

short-term gold price predictions. This indicates that even for gold, which is heavily driven by macroeconomic indicators, the tone of news coverage can enhance predictive models in the immediate term. On the other hand, social media platforms capture the perspectives of individual investors and traders. These tend to be noisier but can sometimes preempt market moves through grassroots signals.

2.1.3 Sentiment Analysis Techniques

The method for extracting sentiment from text has evolved rapidly. Earlier works often relied on lexicon-based methods or simple classifiers. One prominent lexicon-based tool is VADER (Valence Aware Dictionary for sEntiment Reasoning), introduced by Hutto and Gilbert (2014), which is a rule-based model custom to social media language. VADER provides sentiment scores by summing up intensities of sentiment-laden words (accounting for slang, negations, emphasis, etc.), and it was shown to outperform even individual human raters in consistency when classifying tweet sentiments. Due to its efficiency and proven performance on microblog text, VADER has been widely used for quick sentiment gauging on platforms like Twitter or Reddit. However, lexicon methods may miss context or sarcasm, and they often label a lot of financial text as neutral due to domain-specific language. FinBERT, on the other hand, represents the next generation of sentiment analysis for finance. FinBERT is a pre-trained transformer model based on BERT that has been fine-tuned on financial communications and news text (Araci, 2019). It is specifically designed to recognize sentiment in finance-related sentences (e.g. earnings reports, headlines) and has demonstrated superior accuracy on financial sentiment benchmarks. FinBERT effectively handles domain-specific jargon such as "bullish", "hawkish" that general sentiment models might misclassify. The use of FinBERT in forecasting tasks is gaining traction. For example, Zeng and Jiang (2023) integrated FinBERT-based sentiment analysis with an LSTM model for stock trend prediction, finding that incorporating sentiment significantly enhanced the model's ability to anticipate market fluctuations. This study will critically compare VADER and FinBERT outputs on the same dataset to determine which is more suitable for capturing gold-relevant sentiment. It is expected that VADER might excel with casual social media posts, whereas FinBERT may provide more detailed sentiment scoring for news text.

2.1.4   Predictive Modelling Approaches

The choice of modeling technique can greatly affect how sentiment information is utilized. Machine learning ensemble methods like Random Forest (RF) and XGBoost have been popular for tabular financial data due to their ability to handle nonlinear relationships and interactions. Random Forests build multiple decision trees and aggregate their results, offering robustness and a degree of interpretability (feature importance). XGBoost (Extreme Gradient Boosting) is a powerful boosting algorithm known for its high predictive performance and efficiency in many Kaggle competitions; it can capture complex patterns by sequentially improving weaker models. These models, however, do not inherently account for temporal dependencies unless features like lagged values are manually added. In contrast, deep learning methods, particularly Long Short-Term Memory (LSTM) networks, are designed to handle sequence data and remember long-term dependencies in time series. LSTM has been successfully applied to financial time-series forecasting, often outperforming traditional models when sufficient data is available, because it can learn patterns of momentum or mean reversion directly from sequences of past prices and inputs. There is ongoing debate on which approach is superior for financial prediction, and results can vary by context: for example, one comparative study found that LSTM generally outperformed Random Forest in predicting stock price movements, especially for capturing long-term dependencies in the data. On the other hand, another study noted that Random Forest had lower bias and responded faster to sudden price changes compared to LSTM in a stock prediction task. This suggests that LSTM might better capture gradual trends, whereas RF can be more responsive to noise or short-term fluctuations. In this research's context, we will explore both paradigms. The RF and XGBoost models will use engineered features including sentiment scores and possibly technical indicators, providing a straightforward way to gauge the added value of sentiment features via feature importance analysis. The LSTM model will ingest a sequence of past days' data (prices, volumes, sentiments over recent days) to directly forecast the next day's price. Notably, hybrid strategies are also documented. For example, Shi et al. (2022) combined an LSTM with an XGBoost as a post-processor to refine predictions– highlighting that ensemble-of-models or two-stage models can sometimes yield improvements by combining the strengths of different approaches. For feasibility, this

research will focus on comparing standalone models like RF, XGBoost, LSTM under the same data conditions. A critical review of literature suggests that while LSTM may achieve lower error in many cases, tree-based models could rival or outperform deep learning on smaller datasets or more volatile series, and thus all three will be valuable to test.

2.1.5   Evaluation Metrics and Findings

Researchers use various metrics to evaluate prediction accuracy. Common regression metrics include Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which quantify average prediction error (with RMSE penalizing larger errors more). In financial forecasting, beyond these, directional accuracy (the percentage of correct up/down movements predicted) is crucial, since correctly predicting the direction can be as important as the magnitude for trading strategies. Prior studies like Bollen et al. reported high directional accuracy when sentiment was included, as mentioned. Another relevant insight from literature is that sentiment features often shine in improving directional predictions even if the exact numeric error remains moderate. Additionally, some works employ statistical tests to confirm if differences in forecast accuracy are significant. This may be considered in the analysis phase of our study to ensure that any gains from sentiment integration are not due to chance. Finally, the literature emphasizes interpretability where one can ascertain how much a sentiment feature influenced a particular prediction using suitable statistical test. For instance, Hajek and Novotny's fuzzy model inherently provided interpretable rules, revealing conditions under which news sentiment impacted gold price direction. This research remains conscious of the need not only to build an accurate model but also to derive insights. For example, identifying if a surge in negative news sentiment reliably precedes a rise in gold price. The literature review thus supports the research by highlighting the importance of multi-source sentiment, the advancements in NLP for finance, and the comparative strengths of various modeling approaches, all of which inform the design of our methodology.

CHAPTER III

METHODOLOGY

3.1   DATA COLLECTION

The dataset will consist of two main components which is historical gold price data and textual data from news and social media. For gold prices, daily historical closing price of gold per ounce, in USD will be obtained from a reliable source, Yahoo Finance API where it covers a multi-year period from 2015 – 2024. This provides the dependent variable to predict. Corresponding to this timeline, we will gather financial news headlines that pertain to gold or the broader economic and financial context. Potential sources include Reuters or Bloomberg news feeds, and we may leverage an aggregator or an open dataset such as Kaggle's commodity news datasets. Each news item will have a publication date, and we will use the headline or a summary for analysis, assuming that the tone of news on a given day may influence or reflect that day's market sentiment. In addition, social media posts will be collected, focusing on content around gold or market sentiment. The Twitter API can be used to fetch tweets containing keywords like "gold price", "gold investment", or hashtags such as #gold, #XAU (gold's trading symbol), etc. Alternatively, platforms like Reddit could serve as sources. To manage scope, we might limit social data to Twitter. We will collect tweets over the same period as the news, ensuring us to record the date/time and content. Privacy and ethical considerations will be observed by using only publicly available data and anonymizing it as needed.

3.2   DATA PREPROCESSING

Both numeric and text data will be preprocessed for quality. The price series may be adjusted for any missing values or outliers. News headlines and tweets will undergo data cleaning processes such as lowercasing, removal of URLs, hashtags, and other non-

textual noise. We will also group all the text data by day, so it matches the dates. Each news headline on a certain day can also be analyzed its sentiment and perhaps an average or dominant sentiment computed for that day. This results in daily sentiment time series that will be merged with daily prices. It is important to align timing: news published after market hours might influence next day price, but for this study, a simplifying assumption is that daily data aligns with same-day closing prices.

3.3   SENTIMENT ANALYSIS

We will apply VADER and FinBERT sentiment analysis techniques on the textual data. For VADER, we will use the NLTK or vaderSentiment library in Python, which outputs a sentiment score in the range of -1 to +1 for each text input, along with compound sentiment and categorical labels. VADER is well-suited for short social media texts, and its rule-based nature ensures it can handle slang or punctuation emphasis common in tweets. For FinBERT, we will use a pre-trained FinBERT model to infer sentiment on financial texts. FinBERT typically classifies text into positive, negative, or neutral sentiment specific to finance domain terminology. Each news headline or tweet will be fed into FinBERT to get a probabilistic sentiment classification or a sentiment score. The outcome will be two sets of daily sentiment features from VADER and FinBERT respectively. If needed, separate sentiment series for news and for social media will be maintained to later examine their individual effects. We will also conduct a comparative check for example, do VADER and FinBERT agree on the sentiment trend? It is expected FinBERT might flag subtle financial negatives like "downgrade" or "recession fears" that VADER might misclassify as neutral due to domain context. These sentiment scores will then be merged into the modeling dataset by date.

3.4   FEATURE ENGINEERING

In addition to sentiment features, the modeling dataset will include relevant explanatory variables which is prior gold prices and possibly other market data if available. However, to isolate the impact of sentiment, the core comparison will be between a baseline model using only historical prices vs. an enhanced model using prices and sentiment features. We will create lag features such as the previous day's price return, or a moving average of gold price, since models like RF and XGBoost benefit from

explicit features capturing momentum or mean reversion. The inclusion of lagged sentiment might also be tested, under the hypothesis that sentiment could lead price changes with a short delay. All features will be scaled appropriately.

3.5   PREDICTIVE MODELLING

We will develop and train three primary models for one-day-ahead gold price prediction

3.5.1   Random Forest

Using scikit-learn's implementation, with number of trees as estimators tuned and depth limits to prevent overfitting. RF will take in features such as recent price changes and sentiment scores. It will output a regression prediction for the next day's price. We will tune hyperparameters such as tree depth, number of trees, and leaf samples via cross-validation.

3.5.2   XGBoost Regressor

Using the XGBoost library, similarly providing it the feature set. Key hyperparameters to tune include the learning rate, max depth, and number of estimators. XGBoost strength is handling feature interactions and skewed distributions where it might pick up nonlinear relations between sentiment and price that linear models would miss.
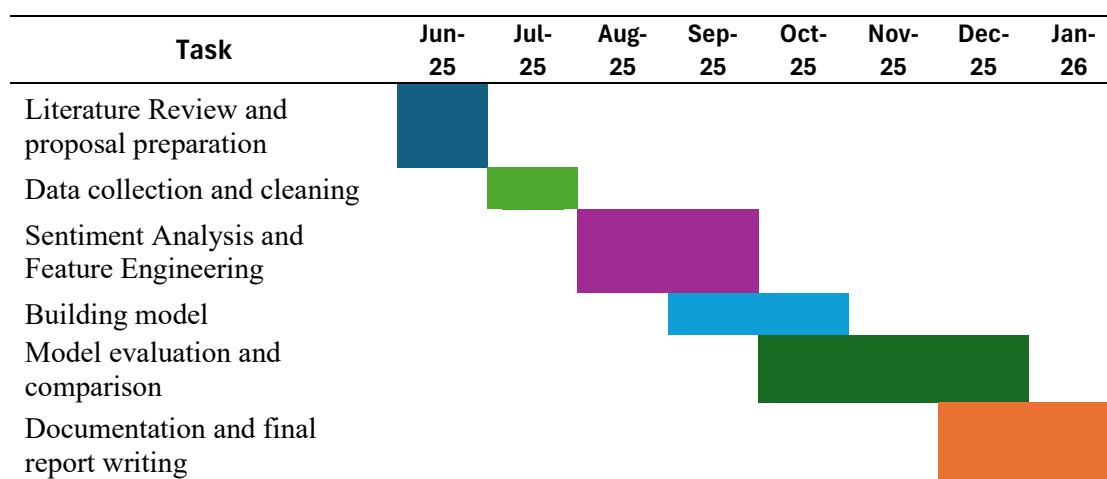
3.5.3   LSTM Neural Network

Libraries such TensorFlow, Keras or PyTorch can be used to build a sequential model. The data will be arranged as sequences of the past 7 or 14 days, with each day's inputs being price change, sentiment scores, etc. The LSTM layer will have a certain number of units to be tuned and use appropriate regularization to prevent overfitting. We will likely include LSTM layer followed by a dense layer to output the next day price. The model will be trained using a training set and validated on a dev set for tuning epochs to avoid overfitting. For time-series integrity, we will make sure the data used for training comes before the data we're trying to predict.

3.6   EVALUATION METRICS

For each model, we will compute Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) on the test set to quantify prediction accuracy in dollar terms. These metrics will tell us how far off the predictions are on average. We will also compute the percentage of days for which the model correctly predicted the direction of price movement compared to the previous day. This metric is important for evaluating the usefulness of the model in a trading context to observe whether the model can get the trend right. We will compare the metrics across models. For instance, does the LSTM yield a lower RMSE than the Random Forest? A particular focus will be on the difference between models that use sentiment features and those that do not. To quantify the worthiness of sentiment, we can train a baseline LSTM which has price only and compare it with a sentiment-enriched LSTM, and similarly for RF and XGBoost, then compare their errors. Additionally, for interpretability, we will examine the Random Forest's feature importance scores and XGBoost's feature importance to see where sentiment ranks among predictors. If, for example, news sentiment emerges as one of the top features in the Random Forest, that provides evidence that sentiment has predictive power.

3.7   RESEARCH TIMELINE (GANN CHART)

Task	Jun-25	Jul-25	Aug-25	Sep-25	Oct-25	Nov-25	Dec-25	Jan-26
Literature Review and proposal preparation	■							
Data collection and cleaning		■						
Sentiment Analysis and Feature Engineering			■					
Building model				■				
Model evaluation and comparison					■	■		
Documentation and final report writing							■	■

REFERENCES

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return: Evidence from the Chinese market. *Knowledge-Based Systems, 69*, 14–23. https://doi.org/10.1016/j.knosys.2014.04.022

Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. DOI: https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). DOI: https://doi.org/10.1145/2939672.2939785

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. DOI: https://doi.org/10.1016/j.ejor.2017.11.054

Hajek, P., & Novotny, V. (2022). Interpretable fuzzy rule-based framework combining financial variables with news sentiment for gold price forecasting. Expert Systems with Applications, 192, 116487. DOI: https://doi.org/10.1016/j.eswa.2021.116487

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Araci, D. (2019). *FinBERT: Financial Sentiment Analysis with Pre-trained Language Models*. arXiv preprint arXiv:1908.10063. DOI: https://doi.org/10.48550/arXiv.1908.10063

Zeng, Q., & Jiang, T. (2023). Financial Sentiment Analysis Using FinBERT with Application in Predicting Stock Movement. arXiv preprint arXiv:2306.02136.

Shi, Y., et al. (2022). Attention-Based CNN-LSTM and XGBoost Hybrid Model for Stock Prediction. (Referenced in Zeng & Jiang, 2023 as a comparative approach).

Zhou, J., & Mengoni, P. (2020). Spot Gold Price Prediction Using Financial News Sentiment Analysis. In 2020 IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp. 758–763.