

# Impact of News on the Commodity Market: Dataset and Results

Ankur Sinha, Tanmay Khandait

Production and Quantitative Methods  
Indian Institute of Management Ahmedabad  
Ahmedabad, India 380015  
asinha@iima.ac.in, tanmayk@iima.ac.in

**Abstract.** Over the last few years, machine learning based methods have been applied to extract information from news flow in the financial domain. However, this information has mostly been in the form of the financial sentiments contained in the news headlines, primarily for the stock prices. In our current work, we propose that various other dimensions of information can be extracted from news headlines, which will be of interest to investors, policy-makers and other practitioners. We propose a framework that extracts information such as past movements and expected directionality in prices, asset comparison and other general information that the news is referring to. We apply this framework to the commodity “Gold” and train the machine learning models using a dataset of 11,412 human-annotated news headlines (released with this study), collected from the period 2000-2019. We experiment to validate the causal effect of news flow on gold prices and observe that the information produced from our framework significantly impacts the future gold price.

**Keywords:** Machine Learning, Natural Language Processing, Text Mining

## Introduction

News has always been one of the primary sources of influence while making financial decisions. With technology emerging, we have news flow across all domains in amounts that exceed the cognitive capacity of an individual, especially in the field of finance. Recent developments in text mining have enabled investors and policy-makers to capitalize on the unstructured information, primarily textual data, for making decisions.

The Semi-Strong Efficient Market Hypothesis [17], has inspired much research on establishing the relationship between the opinions of widely available information and returns of stock and commodity prices [30,14]. News, in the form of headlines [8,13], opinions [30] and various surveys have been known to influence the decision making process of the investors significantly. However, the effect of news items in the context of commodities has not received as wide attention as stocks [27,29,18,19]. It is known that the prices of the commodities are highly

volatile, but very few papers have studied the impact of news on commodity prices, which has the potential to explain some of the uncertainties. In this paper, we bridge this gap by applying news analytics to one of the most important commodities, which is gold. The aim of this study is to provide the investors and policy-makers with a tool to process and analyze large amount of gold-related news that is flowing continuously from multiple geographies.

The works related to public opinions impacting gold commodity prices have emerged in two forms. The first category of papers investigated the effect of various macroeconomic announcements on the price of commodities. Frankel and Hardouvelis [11] were one of the first to study the effect of the announcements on various commodities. Barnhart [2] worked on the effect of money supply announcements and macroeconomic indicators on the commodity prices. Macroeconomic announcements were used to study the effect of these on intra-day gold and silver futures [6] and the volatility of gold market [3]. Some other work that was built up around similar ideas are [4,9,24].

The second category of papers investigated the effect of various news from publishing houses, and opinions and sentiments of people from micro-blogging sites, like twitter, on the price of the gold commodity. To our best knowledge, Mao et al. [20] were the first who considered data from various sources and investigated its effect on the gold prices. Rao et al. [23] worked on similar research using opinions from twitter. Smales [28], was the first to use news sentiments from the Thompson Reuters News Analytics (TRNA) and studied its effect on the gold futures. Other works which used news sentiments in the context of commodities are [10,26]. However, we believe that lack of datasets and dependency on softwares (like TRNA, RavenPack) has hindered the research focused on studying the impact of news on commodities in general and gold in particular. To the best of our knowledge, there are no publicly available datasets that can be used to create machine learning models to extract useful information from commodity news.

Research in the field of Natural Language Processing (NLP) has been focused on both the representation of the text as well as the classification model to process these representations. The development of context-aware pre-trained word embeddings, such as the Bidirectional Encoder Representations from Transformers (BERT) [7], which have been adapted to the domain of finance has shown to outperform the previous best performing models [1]. The previously best performing algorithms have utilized deep-network architectures such as standard Recurrent Neural Network (RNN) [25], Long Short-Term Memory (LSTM) [15], and Gated Recurrent Units (GRU)[5] with pretrained words embeddings such as the Global Vectors for Word Representation (GloVe) [22,21,12].

Through this paper, we introduce a human-annotated dataset of 11,412 news headlines about the gold commodity classified into nine dimensions; i.e., whether the news headline is about commodity prices, asset comparisons, or some other general information; whether the news headline is about the past price movements or future prices movements; and what is the directionality of price movements that the news suggests. In this study, we have compared the performance of various word-embeddings (Tf-Idf, GloVe and BERT) along with various machine

learning algorithms; like Support Vector Machines (SVM), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), and Gated Recurrent Units (GRU); to classify the news headlines into various dimensions. Towards the end of the paper, we perform a causality analysis, which suggests that there is a significant causal relationship between the discussions in the news and the commodity prices. In fact, an important observation that we make is that the impact of news is observable on the gold prices even 24 hours later.

## System Design

In this section, we provide the dataset creation and annotation process, followed by the learning aspects of the system. Figure 1 is an overview of the system that we have designed in this study. A news headline that is received in real-time is fed into our system, which first converts the headline into a vectorized representation. The vectorized representation thereafter enters a classifier model, which classifies the news into one or more of the nine classes. The classes have been identified based on our interactions with the users of the system. The most important step in the system design is the creation of the classifier model for which we needed an annotated dataset.

### Gold News Dataset

With this study, we release a collection of 11,412 human-annotated news headlines dataset from around the world in the period 2000 to 2019, which are specifically about the gold commodity. This dataset was built by scraping news items from various financial news provider sites (Reuters, The Hindu, The Economic Times, Bloomberg etc.) and aggregator sites (Kitco, MetalsDaily etc.). The process of annotation required two crucial tasks; first, deciding the categories in which the news headlines must be classified into; and second, deciding the manual process of annotation.

**Categories for Annotation** Every news item in the headline either mentions movement in prices or other general information related to the gold commodity. Hence, each of these news items were initially classified into price related news, or general news. Consider Ex. 1 that belongs to the price category.

Example 1: *Dec. gold settles at \$1,293.80/oz, up \$8.80, or 0.7%.*

If the news talks about price, we further classify each news headline separately on three dimensions, that tells us if the news item belongs to a particular price movement category or not. The three dimensions that talk about the price movement are:

1. Price Up: This category represents the news headlines indicating if the news discusses the price heading up (irrespective of the movement being in the past or future).

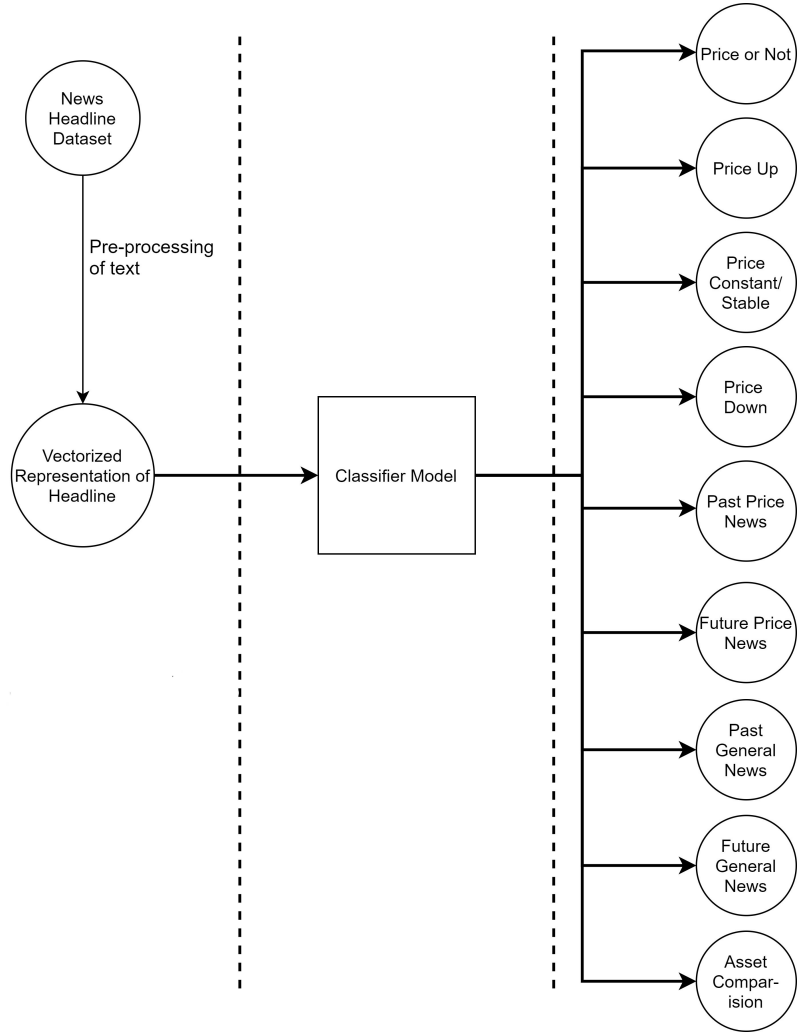


Fig. 1: Overview of System Design

2. Price Constant: This category represents the news headlines indicating if the price has remained constant or stable (irrespective of the movement being in the past or future).
3. Price Down: This category represents the news headlines indicating if the news discusses the price heading down (irrespective of the movement being in the past or future).

For price related news, we also look at the time period dimension, which describes the time period to which the news headline refers to with reference to the price. Consider the following news item.

Example 2: *Gold prices slide \$14.90, or 1.1%, to \$1,289.80 an ounce*

This news item talks about the decline in prices of the gold commodity in the past. Consider another news item.

Example 3: *Gold prices to trade higher today: Angel Commodities*

This news item mentions that the prices of gold might trade higher. Thus, in order to categorize the news events into categories based on time period, we came up with the following two categories that indicate the time period of the news item with reference to the price.

1. Past Price Information: This category classifies the news headlines based on any past information about gold prices.
2. Future Price Information: This category classifies the news headlines based on any future information about gold prices.

News headlines can also provide general information (apart from prices) about imports, exports, production etc. Such news items were also divided based on the time period aspect, but in a different category. Consider the following news item.

Example 4: *Gold imports dip 8% to \$31.72 bn in 2015-16.*

This headline tells us that gold imports have dipped. Since this news item gives us information about the gold commodity other than the price, we need to indicate that the news item talks about past information about the gold commodity, but other than the prices. Consider another news item.

Example 5: *WGC to form panel for setting up spot gold exchange in India.*

This news item also does not refer to gold commodity prices, but it gives us information about an event that is going to happen in the future. Hence, we need to indicate that the news item talks about future information about the gold commodity, but other than its prices. Such future events, though not explicitly talking about gold prices, are often useful for investors and policy makers. Therefore, the two classes that talk about the time period that could be highlighted in the news headlines are:

1. Past General Information: This category classifies the news headlines based on any past information other than the gold prices.
2. Future General Information: This category classifies the news headlines based on any future information other than the gold prices.

Various news headlines compare the movement of prices of two assets. We believe that capturing this information could help in gaining insights into the relationship between the gold commodity and other assets. Hence, we introduce an additional class to indicate that a news headline talks about a comparison purely in the context of the gold commodity with another asset. Consider the following news headline.

Example 6: *Gold notches a gain for a second day as strong dollar pauses its climb.*

The news headline in Ex. 6 indicates that the gold commodity has gained while the dollar has paused its climb.

The annotations of these example headlines are shown in Table 1. It is important to note that a 1 corresponding to a particular news item signifies that it belongs to that specific category, while a 0 signifies otherwise. A news item can also belong to multiple categories (like news item in 3<sup>rd</sup> row in Table 1).

Table 1: Annotation of Examples news headlines into various categories. These news headlines were taken from the dataset.

Sr. No.	News Item	Price or Not	Price Up	Price Const/Price Stable	Price Down	Past Price Info.	Future Price Info.	Past Gen. Info.	Future Gen. Info.	Asset Comp.
1	Dec. gold settles at \$1,293.80/oz, up \$8.80, or 0.7%	1	1	0	0	1	0	0	0	0
2	Feb. gold settles at \$1,282.30/oz, up \$5.60, or 0.4%.	1	1	0	0	1	0	0	0	0
3	Gold ends at a more than 1-week low, but notches slight monthly gain.	1	1	0	1	1	0	0	0	0
4	Gold prices slide \$14.90, or 1.1%, to \$1,289.80 an ounce	1	0	0	1	1	0	0	0	0
5	Gold prices to trade higher today: Angel Commodities	1	1	0	0	0	1	0	0	0
6	Gold imports dip 8% to \$31.72 bn in 2015-16.	0	0	0	0	0	0	1	0	0
7	WGC to form panel for setting up spot gold exchange in India	0	0	0	0	0	0	0	1	0
8	Gold notches a gain for a second day as strong dollar pauses its climb.	1	1	0	0	1	0	0	0	1

**Process of Annotation** Our dataset was manually annotated by three human annotators who were matter experts and were given the following guidelines.

1. All annotations must be done by looking at the news headlines only. No other source like the sub-headlines, news text, etc., must be referred to.
2. All the annotators should independently annotate the headlines without bringing in any inherent bias that could occur from their personal views. Following the above, we arrived at three different series for the annotators. For all the cases where there was discrepancy among the annotators, a consensus-based approach was used to resolve the issue which gave us the fourth series which we refer to as the consensus series. The consensus series has been used in the paper to conduct the experiments.

Table 2 represents the distribution of news headlines into various categories. We also report the inter-annotator’s agreement score using the Cohen’s Kappa statistic measure for every category. The agreement between the annotators was observed to be above 0.85 for all the categories.

Table 2: This table represents the number of news from the consensus series that falls into each category (The total number of items are 11,412). We report the inter-annotator agreement using the Cohen’s Kappa for all the categories.

Aspects Dimensions		True	False	Cohen’s Kappa
Price Related News	Price or Other	9735	1677	0.947
	Price Up	4747	6665	0.882
	Price Constant / Stable	523	10889	0.895
	Price Down	4230	7182	0.902
	Past Price Information	9355	2057	0.912
	Future Price Information	381	11031	0.985
Other News	Other Past Information	1598	9814	0.915
	Other Future Information	82	11330	0.954
	Asset Comparison	2150	9262	0.987

### Classifier Models

Our task is to map a text input to a binary output indicating if the news headline belongs to a particular category or not. Figure 1 is a representation of the system design that we have followed in order to classify the various news headlines. After cleaning, processing and vectorizing the textual data, various classification models then take in the vectorized representations of words and give an output. Every news item is classified into various categories as shown in Figure 1. In our experiments we have used multiple vectorization methods as well as different classifiers. For vectorization, we used word-frequency and word-embedding based ideas, while as classifiers we have used Support Vector Machines (SVM), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM), and Gated Recurrent Units (GRU).

In order to compare and assess the performance of every model with each other, we define a baseline model. The baseline model uses the Term Frequency - Inverse Document Frequency (TF-IDF) weighting scheme to vectorize text and a Support Vector Machine (SVM) classifier to classify these texts into respective categories. The other models use the GloVe word-embeddings to vectorize the texts along with the other classifier algorithms and BERT approach adapted to the financial domain. The performance of every other model is compared with the

baseline model. We also include the performance of pre-trained BERT approach adapted to the financial domain in our study.

The news headline text required pre-processing before vectorization. We followed the following steps for pre-processing:

1. Removal of Punctuation Marks: All kinds of punctuation mark and other special characters were removed from news headlines.
2. Removal of Numbers: All numbers in the news headlines were replaced with a token “NUM”.
3. Changing of Cases: All the news headlines were converted to lower case characters.
4. Filtering of Stop-Words: For the baseline model, where TF-IDF weighting scheme was used to vectorize the text, some stop words were retained while others were removed. Various stop words like *up, down, above, below, under etc.*, were crucial to determine the directionality in the news headlines. Other stop words like *after, before, etc.* were crucial in determining the time in the news headlines. Hence, only specific stop-words like *a, an, the, of, etc.*, were filtered. For the GloVe text vectorization, no stop words were filtered.

Text vectorization is a way to represent textual data quantitatively. We have  $N$  news headlines that were scraped and pre-processed as mentioned above. Using the TF-IDF approach for the baseline model, the entire dataset of  $N$  news headlines is represented by a  $N \times M$  dimensional sparse matrix. The size of  $M$  varies with the consideration of uni-gram, uni-gram—bi-gram and uni-gram—bi-gram—tri-gram tokens. This process is explained in Figure 2a.

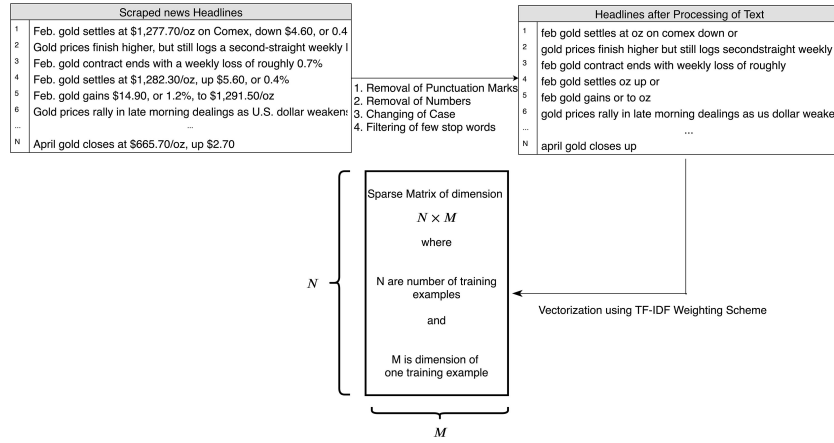
The GloVe pre-trained word-embeddings are known to capture the meaning of a word through a high dimensional vector [22]. For this research, we used the 300-dimensional vectors which were trained on 840 billion tokens through the common crawl. The outline of the process is shown in Figure 2b. It is to be noted that the entire text corpus was represented in the form of a three-dimensional matrix with size  $N \times P \times M$ .

To classify our dataset into various categories, we train different classifiers corresponding to each of the categories present in the dataset. For each category, the SVM classifier classified the news headline into two classes, i.e., whether it belonged to a specific category or not. The other models that were used to compare against our baseline model are some of the more recently developed, sophisticated models. We use the GloVe algorithm to get the word-embeddings for each news headline, which was passed on to RNN, LSTM, and GRU. The Simple and bidirectional versions of these algorithms and the BERT approach adapted to financial domain led to six different classifiers apart from the baseline classifier.

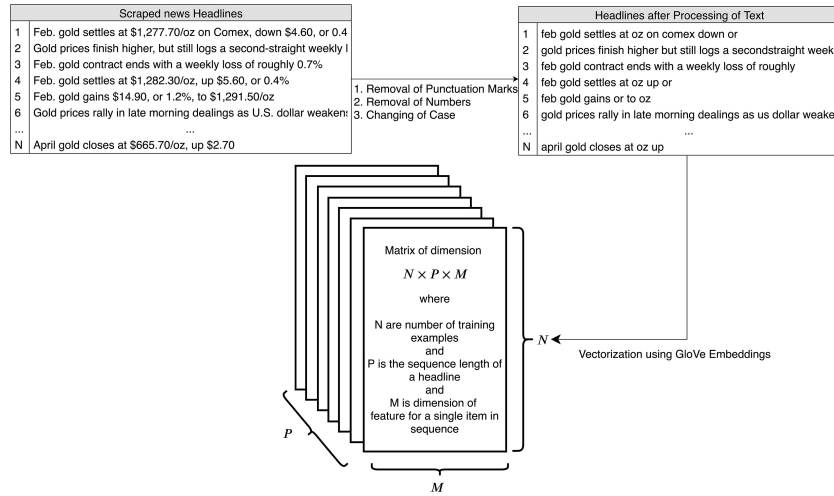
## Results

In this section, we evaluate the performance of various models on categories that are related to gold prices and asset comparison. Table 3 summarizes the





(a) Preparation of input using TF-IDF Model



(b) Preparation of input using GloVe word-embeddings

Fig. 2: Preparation of Input

performance of the baseline model against 6 other models that were used to conduct the experiment.

Table 3 shows that the BERT-based approach adapted to the financial domain has the best performance when compared to other models studied in this paper. Interestingly, the results show that the baseline model also works well when compared against the other models on various categories. While comparing the unidirectional models to the bidirectional models, the bidirectional performed slightly better than its counterpart. The LSTM and GRU performed better than the simple RNN, and if we compare LSTM and GRU, the latter emerges as a

Table 3: Precision, Recall and F1 Values on the test dataset. The percentages in brackets refer to the percentage difference between the F1 score of the baseline model and the corresponding model.

Category		Price or Not	Price Up	Price Constant	Price Down	Past Price	News Future	Price News	Asset Comparison	
SVM (Baseline Model)	Precision	0.968	0.902	0.601	0.914	<b>0.969</b>	0.595	0.992		
	Recall	0.962	0.947	0.881	0.951	<b>0.962</b>	0.949	0.995		
	F1-Score	0.965	0.924	0.715	0.932	<b>0.965</b>	0.732	0.994		
Simple RNN	Precision	0.956	0.891	0.523	0.926	0.965	0.672	0.937		
	Recall	0.93	0.927	0.765	0.878	0.926	0.729	0.973		
	F1-Score	0.943	0.908	0.622	0.901	0.945	0.699	0.955		
RNN		(-2.283%)	(-0.017%)	(-0.13%)	(-0.033%)	(-2.117%)	(-4.444%)	(-0.039%)		
	Bidirectional RNN	Precision	0.971	0.9	0.732	0.921	0.963	0.555	0.965	
		Recall	0.955	0.935	0.712	0.929	0.938	0.934	0.968	
F1-Score		0.963	0.917	0.722	0.925	0.951	0.696	0.966		
LSTM		(-0.199%)	(-0.758%)	(0.979%)	(-0.751%)	(-1.541%)	(-4.869%)	(-2.817%)		
	Simple LSTM	Precision	0.958	0.917	0.711	0.913	0.959	0.656	0.979	
		Recall	0.971	0.924	0.757	0.928	0.949	0.785	0.989	
F1-Score		0.964	0.921	0.734	0.921	0.955	0.715	0.984		
LSTM		(-0.06%)	(-0.325%)	(2.657%)	(-1.18%)	(-1.121%)	(-2.298%)	(-1.006%)		
	Bidirectional LSTM	Precision	0.96	0.927	0.698	0.929	0.964	0.672	0.958	
		Recall	0.973	0.921	0.765	0.917	0.950	0.835	0.995	
F1-Score		0.966	0.924	0.73	0.923	0.957	0.745	0.976		
GRU		(0.154%)	(0%)	(2.098%)	(-0.966%)	(-0.836%)	(1.761%)	(-1.811%)		
	Simple GRU	Precision	0.962	0.909	0.678	0.925	0.958	0.672	0.99	
		Recall	0.972	0.934	0.789	0.931	0.964	0.851	0.995	
F1-Score		0.967	0.921	0.729	0.928	0.961	0.751	0.993		
GRU		(0.266%)	(-0.325%)	(1.958%)	(-0.429%)	(-0.439%)	(2.649%)	(-0.101%)		
	Bidirectional GRU	Precision	<b>0.959</b>	0.924	0.718	0.935	0.967	0.625	0.99	
		Recall	<b>0.976</b>	0.929	0.836	0.916	0.948	0.899	0.99	
F1-Score		<b>0.967</b>	0.927	0.773	0.926	0.958	0.737	0.99		
BERT		(0.273%)	(0.325%)	(8.112%)	(-0.644%)	(-0.806%)	(0.768%)	(-0.402%)		
	BERT	Precision	0.952	<b>0.939</b>	<b>0.966</b>	<b>0.953</b>	0.946	<b>0.983</b>	<b>0.997</b>	
		Recall	0.95	<b>0.939</b>	<b>0.942</b>	<b>0.952</b>	0.944	<b>0.983</b>	<b>0.996</b>	
F1-Score		0.95	<b>0.939</b>	<b>0.95</b>	<b>0.952</b>	0.945	<b>0.983</b>	<b>0.996</b>		
		(-1.554%)	(1.623%)	(32.867%)	(2.146%)	(-2.117%)	(34.29%)	(0.201%)		

winner amongst the various models used. Also the time taken to train GRU was less as compared to LSTM. The categories with a huge class imbalance had a relatively low F1-Score, since we have used no technique to deal with the class imbalance. Many of these results were in line with our expectations and consistent with the literature. The BERT approach was able to better handle the categories with unbalanced classes. In general, we observe that the BERT model performs the best.

## Causality Analysis

This section describes the experiment to establish the causal relationship between the news and the gold prices. We initially provide an overview of the dataset built to conduct this experiment followed by the development of metric based on the daily news items. We then go on to describe the regression model used and present our results.

## Dataset for Evaluation

The task at hand was to establish a relationship between the gold news and gold prices. We establish this relationship for the gold market news and prices in the US from 2017 to 2019. The gold news items were scraped from Kitco, MetalsDaily and Reuters news sites and the prices were obtained for the same period [16].

The news items from the new dataset were classified based on the direction only, i.e., if they talked about price direction up, price direction down or price

direction constant. These were classified using the best model for these three categories (Table 3). We build a metric called the directionality score, which evaluates the overall mood of the news in the context of price movements.

$$\text{Directionality Score } S = \frac{N_{\text{Price Up}} - N_{\text{Price Down}}}{N_{\text{Price Up}} + N_{\text{Price Constant}} + N_{\text{Price Down}}}$$

where  $N$  is the number of news items in the respective categories.

### Model

Using the directionality score, we build a model in order to relate the effect of news items with the gold prices. Let  $S_N$  be the score on day  $N$  and  $P_N$  be the price of gold on day  $N$  at 1700 hrs. The directionality scores are computed using all the news items released between 1700 hours on day  $N - 1$  to 1700 hours on day  $N$ . We study how the change in the directionality score might impact the change in gold prices as follows:

$$S_{N-1} - S_{N-2} \xrightarrow{\text{Predicts}} P_N - P_{N-1} \quad (1)$$

For brevity, we write the Equation 1 as follows.

$$S_{N-1,N-2} \xrightarrow{\text{Predicts}} P_{N,N-1} \quad (2)$$

Using liner regression, this can be expressed as:

$$(P_{N,N-1}) = \alpha + \beta \times (S_{N-1,N-2}) + \epsilon \quad (3)$$

We, therefore, set up the null hypothesis and alternate hypothesis as follows:

**Null Hypothesis ( $H_O$ ):** *There exists no relationship between the directionality score  $S$  and price  $P$ .*

**Alternate Hypothesis ( $H_A$ ):** *There exists a relationship between the directionality score  $S$  and price  $P$ .*

The regression was carried out for two years separately with the first period from April 2017 to March 2018 and the second period from April 2018 to March 2019. For both these periods, we observe that  $\beta$  is significant. The  $p$  - value turns out to be 0.0318 for the first period and 0.00218 for the second period. We therefore reject the null hypothesis and conclude that there exists a causal relationship between the directionality scores  $S$  and the prices  $P$ .

### Conclusion

With this research, we have released a high-quality dataset of 11,412 news headlines about the gold commodity that has been collected from various sources around the world and annotated by human annotators on nine dimensions. This dataset can be used to analyze the various hidden meanings in the news headlines

which might be of interest to investors and policy-makers. In this research, we also studied the performance of various text vectorization methods and classification algorithms in the context of our dataset. Building upon that, we performed a causality analysis, which reveals that the price related news on gold significantly impacts the prices of gold. We believe that this study will open up new avenues for news analytics research in the context of gold and also other commodities, which are known to be highly volatile in prices.

## Acknowledgements

Ankur Sinha would like to acknowledge India Gold Policy Centre (IGPC) for supporting this study under grant number 1815012.

## References

1. Araci, D.: Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063 (2019)
2. Barnhart, S.W.: The effects of macroeconomic announcements on commodity prices. *American Journal of Agricultural Economics* 71(2), 389–403 (1989)
3. Cai, J., Cheung, Y.L., Wong, M.C.: What moves the gold market? *Journal of Futures Markets: Futures, Options, and Other Derivative Products* 21(3), 257–278 (2001)
4. Caporale, G.M., Spagnolo, F., Spagnolo, N.: Macro news and commodity returns. *International Journal of Finance & Economics* 22(1), 68–80 (2017)
5. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
6. Christie-David, R., Chaudhry, M., Koch, T.W.: Do macroeconomics news releases affect gold and silver prices? *Journal of Economics and Business* 52(5), 405–421 (2000)
7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
8. Ederington, L.H., Lee, J.H.: How markets process information: News releases and volatility. *The Journal of Finance* 48(4), 1161–1191 (1993)
9. Elder, J., Miao, H., Ramchander, S.: Impact of macroeconomic news on metal futures. *Journal of Banking & Finance* 36(1), 51–65 (2012)
10. Feuerriegel, S., Neumann, D.: News or noise? how news drives commodity prices. In: ICIS 2013 Proceedings. AIS Electronic Library (2013), 34th International Conference on Information Systems (ICIS 2013); Conference Location: Milan, Italy; Conference Date: December 15-18, 2013
11. Frankel, J.A., Hardouvelis, G.A.: Commodity prices, money surprises and fed credibility. *Journal of Money, Credit and Banking* 17(4), 425–438 (1985)
12. Ghosal, D., Bhatnagar, S., Akhtar, M.S., Ekbal, A., Bhattacharyya, P.: Iitp at semeval-2017 task 5: an ensemble of deep learning and feature based models for financial sentiment analysis. In: Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). pp. 899–903 (2017)

13. Hautsch, N., Groß-Klußmann, A.: When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance* 18, 321–340 (03 2011)
14. Hess, D., Huang, H., Niessen, A.: How do commodity futures respond to macroeconomic news? *Financial Markets and Portfolio Management* 22(2), 127–146 (2008)
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural computation* 9(8), 1735–1780 (1997)
16. MacroTrends: Gold prices - 100 year historical chart. <https://www.macrotrends.net/1333/historical-gold-prices-100-year-chart> (Accessed: 2019-06-15)
17. Malkiel, B.G.: *Efficient Market Hypothesis*, pp. 127–134. Palgrave Macmillan UK, London (1989), [https://doi.org/10.1007/978-1-349-20213-3\\_13](https://doi.org/10.1007/978-1-349-20213-3_13)
18. Malo, P., Sinha, A., Korhonen, P., Wallenius, J., Takala, P.: Good debt or bad debt: Detecting semantic orientations in economic texts. *Journal of the Association for Information Science and Technology* 65(4), 782–796 (2014)
19. Malo, P., Sinha, A., Takala, P., Ahlgren, O., Lappalainen, I.: Learning the roles of directional expressions and domain concepts in financial news analysis. In: 2013 IEEE 13th International Conference on Data Mining Workshops. pp. 945–954. IEEE (2013)
20. Mao, H., Counts, S., Bollen, J.: Predicting financial markets: Comparing survey, news, twitter and search engine data. arXiv preprint arXiv:1112.1051 (2011)
21. Moore, A., Rayson, P.: Lancaster a at semeval-2017 task 5: Evaluation metrics matter: predicting sentiment from financial news headlines. arXiv preprint arXiv:1705.00571 (2017)
22. Pennington, J., Socher, R., Manning, C.D.: Glove: Global vectors for word representation. In: *Empirical Methods in Natural Language Processing (EMNLP)*. pp. 1532–1543 (2014), <http://www.aclweb.org/anthology/D14-1162>
23. Rao, T., Srivastava, S.: Modeling movements in oil, gold, forex and market indices using search volume index and twitter sentiments. In: *Proceedings of the 5th Annual ACM Web Science Conference*. pp. 336–345. ACM (2013)
24. Roache, S.K.: *The Effects of Economic News on Commodity Prices: Is Gold Just Another Commodity?* No. 9-140, International Monetary Fund (2009)
25. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. *nature* 323(6088), 533–536 (1986)
26. Shen, J., Najand, M., Dong, F., He, W.: News and social media emotions in the commodity market. *Review of Behavioral Finance* 9(2), 148–168 (2017)
27. Sinha, A., Kedas, S., Kumar, R., Malo, P.: Buy, sell or hold: entity-aware classification of business news (2019)
28. Smales, L.A.: News sentiment in the gold futures market. *Journal of Banking & Finance* 49, 275–286 (2014)
29. Takala, P., Malo, P., Sinha, A., Ahlgren, O.: Gold-standard for topic-specific sentiment analysis of economic texts. Citeseer
30. Tetlock, P.C.: Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139–1168 (2007)