Using machine learning to analyze the impact of coronavirus pandemic news on the stock markets in GCC countries

Alanoud Al-Maadid [a], Saleh Alhazbi [b,*], Khaled Al-Thelaya [c]

[a] *College of Economics and Finance, Qatar University, Doha, Qatar*
[b] *Department of Computer Science and Engineering, Qatar University, Doha, Qatar*
[c] *College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar*

A B S T R A C T

COVID-19 has resulted in high volatility in financial markets across the world. The goal of this study is to investigate the impact of COVID-19-related news on the stock markets in Gulf Cooperation Council (GCC) countries. The study utilizes machine learning approaches to assess the role of COVID-19 news in stock return predictability in these markets. The results reveal that the stock markets in the United Arab Emirates (UAE), Qatar, Saudi Arabia, and Oman were impacted by coronavirus-related news; however, this news had no impact on the stocks in Bahrain. Moreover, the results indicate that the impacted markets were influenced differently in terms of the quantities and types of news.

1. Introduction

The extremely unpredictable spread of the coronavirus (COVID-19) has led to unprecedented global economic decline. Global economic activity has been heavily influenced by several inevitable reductions in travel, trade, and production. The outbreak of this pandemic has resulted in high volatility in financial markets across the world. Baek et al. (2020) reported that the COVID-19 pandemic impacted stock markets more powerfully than any other infectious disease outbreak that came before. Goodell (2020) presented a comparative study of the impacts of COVID-19 and other natural disasters on the global economy, and the study indicated that COVID-19 is unique in terms of its global scope, wide range and long-term economic damage. In the U.S. stock market, the observed volatility levels surpassed some of the highest volatility levels that occurred in late 1929, the early 1930s, October 1987, and December 2008. Between the last week of Feb. 2020 and the last week of March, the S&P 500 index lost 33% of its value. Similarly, financial markets in South Korea, Australia, Hong Kong, and China dropped by more than 5% daily (Salisu and Vo, 2020), the Nikkei index in Japan plunged more than 30%, and the U.K.'s main index dropped by more than 35% (Zhang et al., 2020).

The analysis of the impact of the pandemic on stock markets is ongoing research, and many studies have adopted different analytical approaches to illustrate and quantify the influence of the pandemic on stock markets from different perspectives. These include the impacts of government measures (Ashraf, 2020a; Zaremba et al., 2020; Haroon and Rizvi, 2020a), spillovers and contagion between markets (Akhtaruzzaman et al., 2020; Okorie and Lin, 2020), the numbers of confirmed infections and deaths (Al-Awadhi et al., 2020; Ashraf, 2020b; Harjoto et al., 2020), and related news (Cepoi, 2020; Ambros et al., 2020; Buckman et al., 2020).

Generally, news is an important factor that greatly impacts stock markets; investors are usually influenced by media reports when

making their decisions. Although news might be incomplete or biased, it impacts investors, especially retail investors, because acquiring and processing real data is costly in terms of time and cognitive efforts (Engelhardt et al., 2020). An extensive literature discusses the impact of news on financial markets (Li, 2018, Sinha, 2016, Beetsma et al., 2013, Smales, 2014, Al-Maadid et al., 2020). This includes different types of news, such as macroeconomic announcements (Chan and Gray, 2018), political events (Burggraf et al., 2020), natural disaster news (Bourdeau-Brien and Kryzanowski, 2017), and global disease-related news (Donadelli et al., 2017). Chan (2003) found that stocks experience a strong negative drift after bad news is reported, while stocks without associated news do not. With the advancement in information technology, investors are becoming exposed to more news through multiple forms of social media, including tweets, social news, and discussion forms, which have a massive influence on the stability of stock markets (Li et al., 2018). However, the volume and contents of news are not only what influence investors but also what attract their attention to this news. Multiple studies, including those of Ekinci and Bulut (2021); Adachi et al. (2017) and Nguyen et al. (2019), used Google search volume (GSV) as a measure of investors' attention and found that it is associated with stock returns. It has also been found that investors are more sensitive to bad news than good news.

The COVID-19 outbreak has been accompanied by a massive volume of day-to-day news, including reports about the numbers of daily new infections, the number of deaths, and the number of recoveries. This news frenzy has increased investors' sense of uncertainty, which has led to high volatility in financial markets. In this study, we investigate the impact of COVID-19-related news on the stock markets in Gulf Cooperation Council (GCC) countries, which include Bahrain, Kuwait, Oman, Qatar, Saudi Arabia, and the United Arab Emirates (UAE). The economies of these countries are oil-based, as they produce approximately 20% of all the oil in the world. Moreover, these countries invest most of their petrodollar surpluses in the global markets; therefore, the financial markets in these countries are affected by global oil prices as well as global crises (Mensi et al., 2017). Moreover, these markets are affected by the volatility of the U.S. market (Alqahtani et al., 2020). Over the past decade, the GCC stock markets have attracted foreign investors and key financial players because of growing investment opportunities in the Gulf region (Ulussever and Demirer, 2017).

Compared to other emerging financial markets, because the economies of these countries mainly depend on oil as their main export and revenue source, the GCC stock markets were hit by a dual shock from both the COVID-19 pandemic and the oil price collapse. By the end of February 2020, the MSCI GCC index declined by 7.3%, and the situation worsened further in March; for example, on March 9, the total GCC market capitalization declined by 162 billion USD (Bahrini and Filfilan, 2020).

The contribution of this study is fourfold. First, in contrast with previous studies (discussed in the next section) that used traditional regression approaches, this study employs machine learning models to explore the impact of COVID-19-related news on financial markets. Machine learning models are more appropriate for analyzing and evaluating the predictability of stock markets than statistical regression approaches because machine learning models have the capability to capture hidden patterns and to consider the effects of multiple factors that influence stock prices; they have better accuracy because of their ability to capture nonlinear modeling trends, which are dominant in financial markets (Bhattacharjee and Bhattacharja, 2019; Stanković et al., 2015). Second, instead of using a single model, we employ three diverse models and compare their accuracies. We use a decision tree (DT) model, which is a traditional algorithm that forms a binary tree by recursively partitioning a dataset into smaller and smaller subsets until the instances in each subset are homogenous or until a threshold that represents the maximum tree depth is reached. The other two models, a random forest (RF) and extreme gradient boosting (XGB), are based on the ensemble technique, and each is built by combining multiple DTs (called weak learners) to achieve improved accuracy. However, they are constructed differently. An RF is based on the bootstrap aggregation technique (bagging), where its weak learners are trained independently and in parallel, and the result of the model is the average of the outputs of all weak learners. On the other hand, XGB is based on a boosting technique, where its weak learners are trained sequentially, the data instances that are predicted incorrectly by weak learners are spotted, and the learner weights are increased to improve the accuracy of the following learners.

Third, the data used in this study cover wider aspects and longer periods of COVID-19 news. In this study, we consider different types of COVID-19-related news, including negative news on Twitter, business news, Google search volumes, numbers of confirmed local cases, numbers of confirmed global cases, numbers of local deaths, and numbers of global deaths. Additionally, the data in this study cover the period from the beginning of 2020 to the beginning of October 2020. This period includes the situation before any local cases were reported and the period of global stock decline, as well as the period during which the markets recovered. Thus, these data capture the role of COVID-19 news in these different situations, not only during the periods before and during the stocks experienced their worst decline in March 2020, as used by other studies. Third, despite the significant increase in the volatility of the GCC stocks as a result of COVID-19, only the study in Bahrini and Filfilan (2020) investigated the effects of the pandemic on these markets; however, the study only investigated the impacts of confirmed cases and deaths during the period from April 1, 2020, to June 26, 2020. Therefore, our study contributes to filling the gap in the literature regarding the understanding of the impact of COVID-19 on GCC stocks.

Overall, this study finds that COVID-19-related news affected stock markets in the UAE, Qatar, Saudi Arabia, and Oman but had no influence on the stock market in Bahrain. Moreover, the influence of news varied among the impacted markets; the markets in the UAE and Qatar were influenced more than the markets in Saudi Arabia and Oman, which were influenced only slightly.

The rest of this paper has the following structure. Section 2 presents a review of the relevant literature. Section 3 describes the utilized dataset. Section 4 illustrates the methodology and the machine learning models used in this study. The empirical results are presented and discussed in Section 5, while Section 6 concludes the study.

2. Literature review

The literature on the impact of COVID-19 on different aspects of life is evolving. The massive level of media coverage associated

with this pandemic has affected investors' behaviors in stock markets. Akhtaruzzaman et al. (2021) found that the media played the main role in transmitting financial contagion across stock markets. To study this impact, Buckman et al. (2020) used news sentiment analysis. Their study employed the Daily News Sentiment Index (DNSI) in the U.S., exploiting the lexical approach and machine learning to process economics-related news articles and provide real-time data representing news sentiment scores. The news sentiment index was compared with the University of Michigan's index of consumer sentiment. The two indices were found to be correlated and exhibited similar or close levels of response to different economic events. The results showed that the increasing coverage of COVID-19 through mid-March was synchronized with the decline in the news sentiment index. Similarly, Lee (2020) used the DNSI in addition to the Google search volumes of terms related to COVID-19 to study the impact of COVID-19 sentiments on S&P sector indices. The results showed that the DNSI is positively related to all sectors but with distinct effects across the sectors.

The mutual impact between stock markets and the media was also addressed in Mamaysky (2020), where news sentiment analysis was used to explore the mutual relations between COVID-19 news articles from Reuters and the prices of five asset classes, including the S&P and VIX indices. The study ran a linear regression analysis to infer the causal relations between news and financial economic indices. The author concluded that there is a strong mutual relationship between stock prices and news. It was found that news can be used to infer prices in the future and vice versa. Similarly, to study the impact of media coverage on stock market volatility, Haroon and Rizvi (2020b) used a panic index (represented by the Ravenpack finance index) to perform sentiment analysis based on COVID-19 news. They employed exponential generalized autoregressive conditional heteroskedasticity (GARCH) models to study the impact of the media on the price returns of multiple stock market indices in the U.S. based on sectoral index data. However, the study found that media coverage and news sentiment had little to moderate impacts on price volatility. The same conclusion was found by Ambros et al. (2020), who performed a regression analysis to study the impact of news sentiment on the high-frequency price returns of nine stock market indices. Nevertheless, the study acknowledged that COVID-19 news partially influences market uncertainty.

Cepoi (2020) investigated the reactions of stock markets to COVID-19-related news in the six countries with the most infections, including the U.S., the U.K., Germany, France, Spain, and Italy. It was found that the stocks in these countries presented asymmetric dependencies with news related to COVID-19. In the same way, Salisu and Vo (2020) examined how COVID-19 news has influenced the stock markets in the twenty countries hit worst by the pandemic. To select these countries, they relied on local official statistics of reported cases and deaths. The same countries were also selected in a study by Cepoi (2020), along with China, Switzerland, the Netherlands, Belgium, South Korea, Turkey, Austria, Canada, Portugal, Brazil, Norway, Australia, and Sweden. The study evaluated the role of health news collected through Google search trends on the predictability of stocks in these countries during the period from January 1 to March 30, 2020. The study revealed that the use of health news improved stock forecasts and had a statistically significant negative effect on stock returns. Engelhardt et al. (2020) found that stock markets have mainly been driven by news attention during the COVID-19 pandemic. The authors used a sample of 64 national stock markets, representing 94% of the global gross domestic product (GDP), and used Google search volumes to explore the impact of COVID-19 news on these markets. Moreover, the study estimated that the cost of news hype was approximately USD 200 billion.

The study in Kamal et al. (2021) used the event-study methodology to explore the short-term effects of pandemic news on maritime shipping stocks in the NYSE during eight major COVID-19 news events. The authors considered four of these events as pessimistic and four of them as optimistic. The pessimistic events included the identification of the first cases in China, the outbreak of the pandemic in Italy, the announcement of the World Health Organization (WHO) declaring COVID-19 as a global pandemic, and reaching the new record of over 1 million cases in the US. The optimistic events included the news of easing the lockdown restrictions in many countries, the announcement of extra economic stimuli in the US, the announcement of Oxford University regarding the strong immune reaction to the developed vaccine, and the announcement of the first vaccine approval in Russia. The study found that investors exhibited adverse reactions to pessimistic news and positive overreactions to optimistic news. Similarly, Goodell and Huynh (2020) used the event-study methodology to assess the impact of sudden COVID-19-related news on industrial stocks in the US alongside an analysis of the attention levels paid by investors to COVID-19. They found that public attention was associated with abnormal returns in several industries, which were also associated with COVID-related news announcements.

Duan et al. (2021) studied the impact of media sentiments related to COVID-19 on Chinese stocks. The authors collected 1.8 million pieces of official news and 4.5 million Sina Weibo microblogs. They built two sentiment indices by classifying each entry in the dataset as an optimistic, neutral, or pessimistic entry by using a support vector machine (SVM). Then, they used a regression model to investigate the relations between these two indices and Chinese stock market returns and found that the stock returns were positively predicted by the sentiments with both platforms. Baek and Lee (2021) used the daily death rate and recovery rate of COVID-19 as proxies for bad and good news, respectively, and they employed the BEKK-multivariate GARCH model to examine the influences of these rates on the US stock market. The results showed that the stock market was significantly and positively impacted by the volatility of the death rate (bad news) and negatively associated with the volatility of the recovery rate (good news). Furthermore, the study found that the US stock market was affected by bad news much more than by good news.

3. Data

The data used in this study cover the period from the January 1, 2020, to the October 4. 2020, to analyze the impact of COVID-19 news on five GCC market indices. They include the Bahrain All Share Index, Muscat Securities MSM 30 Index, Qatar Exchange Index, Tadawul All Share Index, and Abu Dhabi Securities Market General Index, which represent stock markets in Bahrain, Oman, Qatar, Saudi Arabia, and the UAE, respectively. The Kuwait stock market is excluded because its data are unavailable. The daily data of each market are downloaded from the Investing.com website. The daily closing prices are used to compute the daily returns by setting $\hat{y} =$
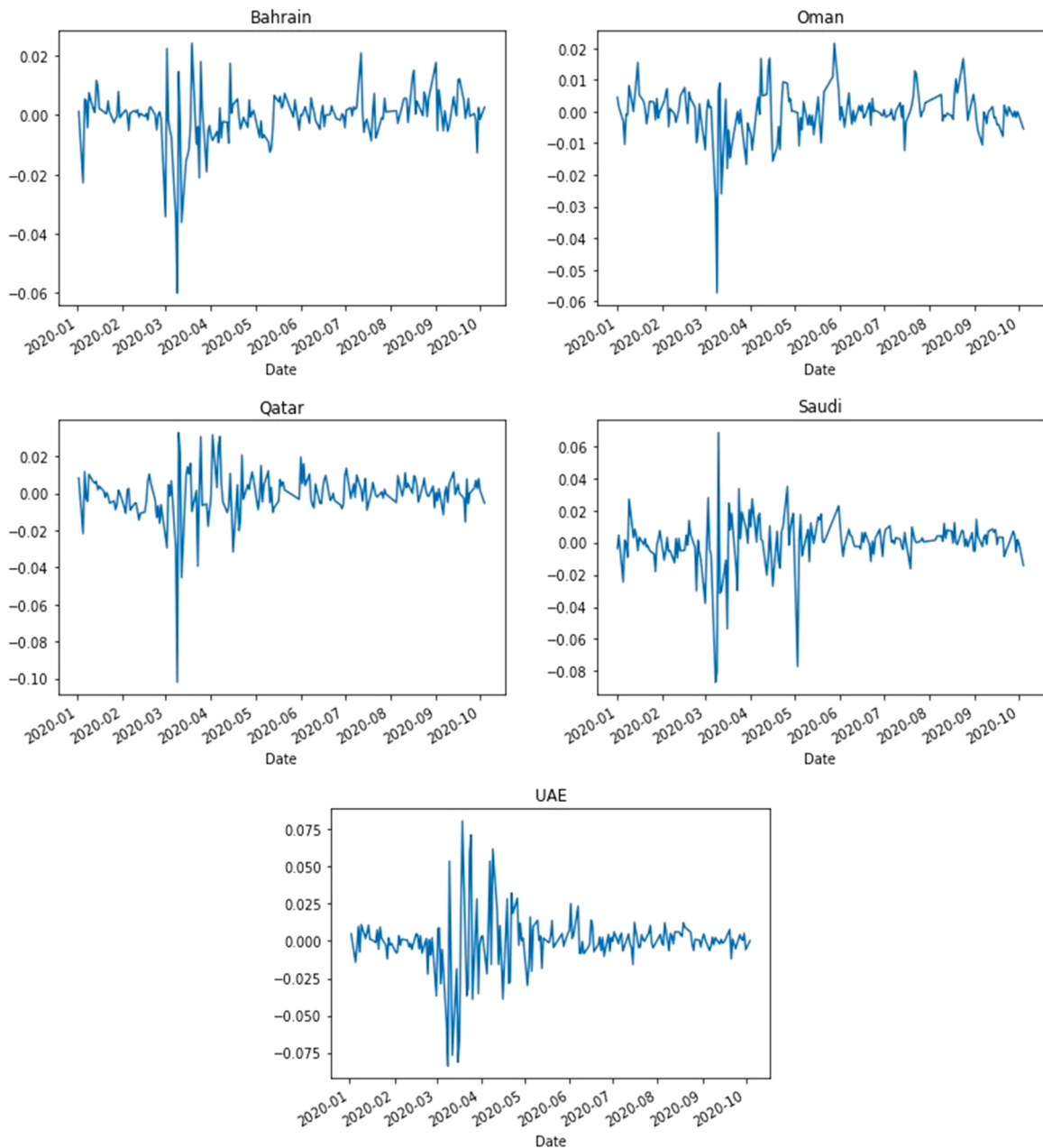
Fig. 1. The returns of the daily closing prices of the five GCC stock market indices.

$\ln\left(\frac{y_i}{y_{i-1}}\right)$. The daily returns of the closing prices of the five GCC stock market indices are illustrated in Fig. 1.

The dataset of each market contains economic and COVID-19-related news indicators. Both data types are obtained from Bloomberg. The economic indicators represent the control variables, including local interest rates, local GDP surprises, i oil prices and U.S. stock volatility index (VIX) data. The GDP surprises are calculated as the differences between the Bloomberg consensus forecasts that are usually released before the actual GDP announcements and the actual GDP releases; this indicator captures the surprise that might be experienced by the local stock market. We include the oil prices and VIX indicators because of their effects on market predictability (Alhazbi et al., 2020). The correlations between the stock markets in GCC countries and oil prices are well documented and exhibit volatility comovement (Alqahtani et al., 2019); the effect of oil prices was significantly larger during COVID-19 than during the previous period (Abuzayed and Al-Fayoumi, 2021). Additionally, it is found that VIX has a significant positive influence on GCC stocks, and U.S. stock uncertainty spreads to the GCC markets (Benlagha and El Omari, 2021). Furthermore, the GCC markets exhibit connectedness (Hung, 2021), so to capture the spillover effects on each market, we also include the daily returns of the other GCC

Table 1
Descriptive statistics of the data for Bahrain market.

	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness	Jarque-Bera	Probability
Total Business News	4.05	6.93	1	61	2	30.4	4.54	6109.68	0
GDP Surprises	-0.34	0.61	-3.5	0.2	-0.23	23.99	-4.54	6055.19	0
Google Search Trends	72.27	44.45	1	184	66	2.4	0.37	10.02	0.01
Interest Rate	2.61	0.84	1.7	4	2.25	2.18	0.91	46.32	0
Negative Business News	1	0	1	1	1	0	0	0.38	0.83
Negative Twitter News	1.16	0.69	1	5	1	27.82	4.98	1102.5	0
Number of Cases - Global	131,230	106,682.7	1	323,485	99,664	1.65	0.27	23.68	0
Number of Cases - Local	337.92	252.28	1	1305	358	3.02	0.44	6.83	0.03
Number of Deaths - Global	3977.95	2502.28	1	10,434	4649	2.23	-0.34	11.36	0
Number of Deaths - Local	2.21	1.39	1	7	2	4.34	1.34	43.99	0
Number of Recoveries - Local	329.15	242.98	1	858	344	1.83	0.1	11.95	0
Oil Prices	37.64	12.42	-37.6	63.27	39.87	8.06	-1.07	276.55	0
Stock Market Daily Returns	0	0.01	-0.06	0.02	0	14.03	-2.02	1063.29	0
Total Twitter News	6.96	4.36	1	22	6	3.68	0.98	36.82	0
VIX Index	30.3	13.66	12.1	82.69	27.68	5.41	1.41	112.98	0

Table 2
Descriptive statistics of the data for Oman market.

	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness	Jarque-Bera	Probability
Total Business News	3.81	5.48	1	29	2	11.36	2.93	764.4	0
GDP Surprises	-0.52	0.43	-1.8	-0.01	-0.5	3.99	-1.12	69.74	0
Google Search Trends	63.1	45.81	1	184	56	2.51	0.64	20.29	0
Interest Rate	0.93	0.71	0.5	2.28	0.5	2.3	1.11	62.77	0
Negative Business News	0	0	0	0	0	0	0	0	0
Negative Twitter News	1.2	0.52	1	3	1	8.39	2.53	45.63	0
Number of Cases - Global	131,230	106,682.7	1	323,485	99,664	1.65	0.27	23.68	0
Number of Cases - Local	557.87	541.16	1	2685	363	3.46	0.99	31.57	0
Number of Deaths - Global	3977.95	2502.28	1	10,434	4649	2.23	-0.34	11.36	0
Number of Deaths - Local	7.25	7.66	1	67	6	30.92	4.27	4583.18	0
Number of Recoveries - Local	610.11	860.19	1	8382	318	46.69	5.33	12,472.15	0
Oil Prices	37.64	12.42	-37.6	63.27	39.87	8.06	-1.07	276.55	0
Stock Market Daily Returns	0	0.01	-0.06	0.02	0	17.06	-2.15	1638.98	0
Total Twitter News	5.96	4.09	1	22	5	4.6	1.16	65.04	0
VIX Index	30.3	13.66	12.1	82.69	27.68	5.41	1.41	112.98	0

Table 3
Descriptive statistics of the data for Qatar market.

	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness	Jarque-Bera	Probability
Total Business News	4.61	5.65	1	37	3	13.81	3.15	1452.65	0
GDP Surprises	-0.2	0.37	-1.05	0.3	-0.1	2.49	-0.51	14.82	0
Google Search Trends	77.72	54.55	1	209	63.5	2.35	0.62	20.97	0
Interest Rate	1.25	0.42	1	2	1	2.41	1.15	46.32	0
Negative Business News	3		3	3	3	0	0	0.38	0.83
Negative Twitter News	1.69	1.49	1	6	1	6.39	2.11	15.9	0
Number of Cases - Global	131,230	106,682.7	1	323,485	99,664	1.65	0.27	23.68	0
Number of Cases - Local	587.7	554.4	1	2355	296	3.06	1.1	43.85	0
Number of Deaths - Global	3977.95	2502.28	1	10,434	4649	2.23	-0.34	11.36	0
Number of Deaths - Local	2.08	1.24	1	7	2	4.51	1.25	36.87	0
Number of Recoveries - Local	626.78	853.04	1	5235	256	12.04	2.61	893.05	0
Oil Prices	37.64	12.42	-37.6	63.27	39.87	8.06	-1.07	276.55	0
Stock Market Daily Returns	0	0.01	-0.1	0.03	0	22.49	-2.69	3168.12	0
Total Twitter News	8.25	7.15	1	65	6.5	21.04	3.08	3391.46	0
VIX Index	30.3	13.66	12.1	82.69	27.68	5.41	1.41	112.98	0

markets in the control variables.

The COVID-19-related news indicators are also obtained from Bloomberg; these indicators include the number of daily news items on Twitter that are related to each country, the daily ratio of negative items to total items, the number of daily articles in newspapers and magazines that discuss coronavirus and business news related to each country, the daily ratio of negative articles to total articles, and daily Google search trend data. Generally, Google search trends are indicators of the relative popularity levels of all searches on Google at a specific time and place. In our study, Google search trends represent an indication of the community's interest in news about the coronavirus pandemic in each country. Additionally, to capture the severity of the health situation in each country and the

Table 4
Descriptive statistics of the data for Saudi market.

	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness	Jarque-Bera	Probability
Total Business News	11.08	12.79	1	89	7	13.8	2.91	1459.72	0
GDP Surprises	-0.26	0.39	-1.6	0.76	-0.3	6.46	-0.61	155.78	0
Google Search Trends	48.03	38.51	2	190	38	4.02	1.14	66.02	0
Interest Rate	1.31	0.53	1	2.25	1	2.36	1.13	64.05	0
Negative Business News	1.67	1.14	1	7	1	12.4	2.79	239.15	0
Negative Twitter News	2.87	2.25	1	10	2	3.95	1.32	49.57	0
Number of Cases - Global	131,230	106,682.7	1	323,485	99,664	1.65	0.27	23.68	0
Number of Cases - Local	1599.99	1238	1	4919	1355	2.56	0.68	17.74	0
Number of Deaths - Global	3977.95	2502.28	1	10,434	4649	2.23	-0.34	11.36	0
Number of Deaths - Local	25.66	14.88	1	58	29	1.88	-0.14	10.49	0.01
Number of Recoveries - Local	1607.43	1351.7	1	7718	1380	4.77	1.11	67.25	0
Oil Prices	37.64	12.42	-37.6	63.27	39.87	8.06	-1.07	276.55	0
Stock Market Daily Returns	0	0.02	-0.09	0.07	0	13.14	-1.76	892.99	0
Total Twitter News	38.56	35	1	246	31.5	7.93	1.69	368.75	0
VIX Index	30.3	13.66	12.1	82.69	27.68	5.41	1.41	112.98	0

Table 5
Descriptive statistics of the data for UAE market.

	Mean	Std. Dev.	Min	Max	Median	Kurtosis	Skewness	Jarque-Bera	Probability
Total Business News	9.69	10.65	1	69	6	9.26	2.25	561.78	0
GDP Surprises	-0.35	0.71	-3.3	0.23	-0.2	14.79	-3.52	2183.07	0
Google Search Trends	70.56	44.76	2	199	60	3.33	0.79	28.36	0
Interest Rate	1.03	0.56	0.6	2	0.75	2.17	1.02	55.81	0
Negative Business News	1.33	0.58	1	2	1	1.5	0.71	0.53	0.77
Negative Twitter News	2.92	1.77	1	8	3	3.17	0.88	21.52	0
Number of Cases - Global	131,230	106,682.7	1	323,485	99,664	1.65	0.27	23.68	0
Number of Cases - Local	464.49	284.15	1	1231	437	2.7	0.34	5.13	0.08
Number of Deaths - Global	3977.95	2502.28	1	10,434	4649	2.23	-0.34	11.36	0
Number of Deaths - Local	2.68	2.36	1	13	2	7.26	2.08	234.95	0
Number of Recoveries - Local	440.62	352.54	1	2443	398	7.08	1.32	197.33	0
Oil Prices	37.64	12.42	-37.6	63.27	39.87	8.06	-1.07	276.55	0
Stock Market Daily Returns	0	0.02	-0.08	0.08	0	9.07	-0.33	296.62	0
Total Twitter News	65.17	46.78	1	209	57.5	2.87	0.55	11.72	0
VIX Index	30.3	13.66	12.1	82.69	27.68	5.41	1.41	112.98	0

global severity, we include the numbers of cases, deaths, and recoveries at both local and global scales. A statistical description of the data is presented in Tables 1, 2, 3, 4, and 5 for Bahrain, Oman, Qatar, Saudi Arabia, and the UAE, respectively.

4. Methodology

In this study, we utilize machine learning approaches to analyze the impact of COVID-19-related news on GCC stocks. We use three types of machine learning models to perform the analysis: decision trees (DTs), an RF, and XGB. To assess the role of COVID-19 news in the predictability of the stock returns in each country, we conduct two experiments by using the three models for each country. In the first experiment, we use only the control variables (economic indicators), which include the interest rate, GDP surprises, VIX, oil price, and daily returns of other GCC markets. In the second experiment, we use the economic indicator data in addition to the news data. The news data include the number of daily news items on Twitter, the daily ratio of negative Twitter news items, the number of daily business news items, the ratio of daily negative business news items, the daily Google search trends, the number of daily confirmed local cases, the number of daily local deaths, the number of daily global cases, the number of global deaths, and the number of daily local recoveries. The data are split into training (80%) and testing (20%) sets, and we use 10-fold cross validation to measure the performance of each model in both experiments.

The difference between the two experiments in terms of the accuracy of each model represents the impact of the utilized news on the model performance, which is measured by using the root mean square error (RMSE). The RMSE measures the discrepancy between the predicted values and the actual values; therefore, smaller RMSE values indicate high model accuracy. The RMSE is calculated as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2} \tag{1}$$

where $\widehat{y}_i$ represents the predicted price returns produced by the regression model and y represents the actual price returns. All the experiments are conducted in Python with the Scikit-learn (Pedregosa et al., 2011) library. The data are normalized between 0 and 1 via min-max normalization to scale the data variability according to the following formula.

$$\widehat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x represents the actual input variable and $\widehat{x}_i$ represents the scaled values. The following subsections describe the three models used in this study.

4.1. Decision tree (DT)

A regression tree is constructed via top-down induction by splitting the given samples based on the input feature set; the optimal split is obtained based on one input feature, and the best gain metric criteria are considered at each step. Ultimately, a set of rules is maintained based on the best metric criteria. This procedure is repeated recursively until splitting does not produce better predictions or the subset of leaves has values similar to those of the target variable. The recursive procedure can also be stopped if the maximum depth limit is reached (Loh, 2011).

Different metrics are commonly used to determine the best subset splits for measuring the homogeneity of the predicted variables for each subset. In our experiments, the MSE is used as the objective function of the splitting criterion, where different subsets are compared, and the best split is selected. For example, to split a particular set into two subsets S_A and S_B, one must seek the plane that minimizes the squared errors of the two subsets, as follows (Loh, 2011).

$$\sum_{i \in S_A} (y_i - \widehat{y}_i)^2 + \sum_{i \in S_B} (y_i - \widehat{y}_i)^2 \tag{3}$$

where y_i and $\widehat{y}_i$ represent the target variable before and after performing the split decision, respectively. Therefore, to determine the level of improvement yielded by the split rule, variance reduction is employed in cases where the target variable is continuous. To measure the variance reduction I_v for any split rule at any node N in the regression tree, the regression decision tree uses the following formula (Loh, 2011).

$$I_v(N) = \frac{1}{|S|^2} \sum_{i \in S} \sum_{j \in S} \frac{1}{2} (y_i - y_j)^2 - \left(\frac{1}{|S_A|^2} \sum_{i \in S_A} \sum_{j \in S_A} \frac{1}{2} (y_i - y_j)^2 + \frac{1}{|S_B|^2} \sum_{i \in S_B} \sum_{j \in S_B} \frac{1}{2} (y_i - y_j)^2 \right) \tag{4}$$

In both experiments, we set three stopping criteria: the maximum depth is five, the minimum number of samples is equal to two for each split, and the minimum number of samples that must be in a leaf is set to one.

4.2. Random forest (RF)

An RF is composed of a collection of DTs trained by using a bootstrap sample drawn with replacement from the training data. Each tree in the forest is constructed based on an input random vector that aims to add a randomness factor to the process to avoid overfitting the data. The collection of tree predictors is denoted by $h(\mathbf{x}; \eta_k), k = 1,.,K$, where $\mathbf{x}$ represents the input data and η_k denotes independent and identically distributed random vectors. For regression problems, the output of an RF is an estimation of a continuous variable Y. The final predictions of the tree predictors are then averaged to produce final prediction values, as follows:

$$h(\mathbf{x}) = \left(\frac{1}{K} \right) \sum_{k=1}^{K} h(\mathbf{x}; \eta_k) \tag{5}$$

The bootstrap aggregation technique is used by the RF algorithm to conduct the training process. For each DT, a sample from the training data is selected with replacement to train the tree. The bootstrapping procedure reduces the variance of the model and does not increase the bias. Unlike a traditional DT, an RF performs a random split on each node based on a random sample of input features. As in the DT mode, we set three stopping criteria: the maximum depth is five, the minimum number of samples is equal to two for each split, and the minimum number of samples that must be in a leaf is set to one. Furthermore, we set the number of trees (estimators) to 100. These hyperparameters are the same in both experiments.

4.3. XGBoost regression (XGB)

XGB is a DT-based method developed based on boosting tree models that uses a gradient boosting technique. This algorithm outperforms many machine learning techniques on small-to-medium structured/tabular data. The difference between this algorithm and other traditional boosting tree models is that it enables parallel CPU computing via multithreading. It also utilizes different methods to prevent overfitting.

Gradient boosting is employed by the XGB algorithm to combine many DT learning models into a stronger learner in an iterative manner. The residual at each iteration is used to improve the prediction accuracy of the previous predictor to optimize the objective loss function. The objective function also includes a regularization term to prevent overfitting. The objective function is given by

$$J(\beta) = L(\beta) + \omega(\beta) \tag{6}$$

where β represents the parameters adjusted by the training data, the loss function is denoted by L, and ω is a regularization term that is

Table 6

The hyperparameters used for the three machine learning models.

Parameters	DT	RF	XGB
Impurity function	Squared Error	Squared Error	Squared Error
Tree depth	5 levels	5 levels	6 levels
Minimum number of samples required to split	2 samples	2 samples	NA
Minimum number of samples required to be at leaf nodes	1 sample	1 sample	NA
The number of estimators	1	100	100
learning rate	NA	NA	0.2
Pseudoregularization (Gamma)	NA	NA	0.08
L1 regularization	NA	NA	1
L2 regularization	NA	NA	1
Gradient boosting	NA	NA	Vanilla gradient

used to prevent overfitting based on model complexity measures. The predicted value $\widehat{y}_i$ at time t is the average of the outputs of a collection of k decision trees (denoted by S), as follows (Zhang et al., 2018)

$$\widehat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \widehat{y}_i^{(t-1)} + f_t(x_i), f_k \in S \tag{7}$$

The objective function J at time t can be defined as

$$J^{(t)} = \sum_{i=1}^{n} L(y_i, \widehat{y}_i) + \sum_{k=1}^{t} \omega(f_k), \tag{8}$$

where n is the number of predictions and $\widehat{y}_i^{(t)}$ is as defined in 7. However, the regularization term $\omega(f_k)$ is defined as

$$\omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{9}$$

where the number of leaves in the tree is denoted by T and the complexity of each node is also used in the regularization function and denoted by γ. The penalty scaling parameter is denoted by λ, and the scores-of-leaves vector is denoted by w. A second-order Taylor expansion is then utilized for the loss function instead of the first-order expansion, which is normally used in the traditional gradient boosting approach. The objective function can be defined using the MSE loss function (Chen and Guestrin, 2016):

$$J^{(t)} \approx \sum_{i=1}^{n} \left[g_i w_{q(x_i)} + \frac{1}{2}\left(h_i w_{q(x_i)}^2\right)\right] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{10}$$

where the first and second derivatives of the MSE loss function are denoted by g_i and h_i, respectively. $q(.)$ is a function that assigns data points to the corresponding leaf. The loss function can also be expressed by the sum of the loss values of all leaf nodes as follows:

$$J^{(t)} \approx \sum_{j=1}^{T} \left[\left(\sum_{i \in I_j} g_i\right)\omega_j + \frac{1}{2}\left(\sum_{i \in I_j} h_i + \lambda\right)\omega_j^2\right] + \gamma T \tag{11}$$

More details about the XGB model can be found in Chen and Guestrin (2016). As in the RF, we set the number of trees in this model to 100, the maximum tree depth is set to six, the learning rate is set to 0.2 in both experiments, and the gamma regularization term is set to 0.08. Table 6 illustrates the hyperparameters used for the three models.

4.4. Feature importance

To identify which indicators in the news dataset have greater impacts on the markets in the UAE, Qatar, and Saudi Arabia, we use the feature importance metric offered by the scikit-learn machine learning library. It is the score that is assigned to each feature and represents the extent to which the feature influences the model relative to other features. In other words, this metric represents the model accuracy reduction induced when the corresponding feature is removed. Since we perform 10-fold cross validation and we have three models for each country, we average the importance of each feature over the thirty trained model iterations for each country.

In scikit-learn, the importance of a feature in a DT is calculated by measuring how effective the feature is at reducing the impurities of nodes, so it is computed as the normalized total of the weighted impurity decreases in all nodes that are split by using that feature. The impurity of a node in a DT for regression is calculated by using the MSE (which is also called variance reduction), as follows:

$$v = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu)^2 \tag{12}$$

where n is the number of instances in a node, y_i is the value of an instance, and μ is the mean of the instances in this node. The weighted

Table 7
Comparing the accuracies of the three models with and without COVID-19 news.

		Bahrain	Oman	Saudi Arabia	Qatar	UAE
DT	Without COVID-19 news	0.1451	0.1215	0.1574	0.1361	0.1599
	With COVID-19 news	0.1475	0.1150	0.1392	0.1160	0.1393
	Difference	-0.0024	0.0065	0.0182	0.0201	0.0206
RF	Without COVID-19 news	0.1209	0.1053	0.1342	0.1193	0.1439
	With COVID-19 news	0.1225	0.1043	0.1240	0.1060	0.1305
	Difference	-0.0016	0.0010	0.0102	0.0133	0.0134
XGB	Without COVID-19 news	0.1132	0.0997	0.1215	0.1092	0.1391
	With COVID-19 news	0.1155	0.0981	0.1139	0.0980	0.1154
	Difference	-0.0023	0.0016	0.0076	0.0112	0.0237

impurity decrease (reduction) of a split in node i based on feature j is calculated as follows:

$$r_{ij} = w_i v_i - w_{left(i)} v_{left(i)} - w_{right(i)} v_{right(i)} \tag{13}$$

where the weight is the ratio of the number of instances in a node to the total number of instances before splitting. "Left" and "right" represent the left and right children of a node in the tree, respectively. The normalized importance of feature j is calculated by dividing the total variance reduction achieved when the feature is used in all nodes by the total variance reduction achieved by using all features as follows:

$$f_j = \frac{\sum_i r_{ij}}{\sum_k \sum_n r_{kn}} \tag{14}$$

The feature importance for the RF and XGB models is calculated by dividing the total feature importance in all the trees in the model and the number of trees in the model.

5. Results and discussion

Table 7 and Fig. 2 present the results of the two experiments conducted for each country with each of the three models. The results show that the XGB model has the highest accuracies in both experiments for all stock markets, and its RMSEs are the smallest in both experiments compared to those of the other two models. The RF model ranks second in terms of accuracy for all markets in both experiments, and the DT model has the lowest accuracy.

The results indicate that there are differences between the GCC countries in terms of the impact of coronavirus news on model predictability. The accuracies of the three models are improved when COVID-19 news indicators are included among the input features for market predictions in the UAE, Qatar, Saudi Arabia, and Oman. However, including these indicators does not improve the accuracy of any of these models in the case of the Bahrain stock market.

Moreover, the improvements in the prediction abilities of the three models vary among the impacted markets; the improvement in the Oman market case is very small compared to those in the UAE, Saudi Arabia, and Qatar market cases. Furthermore, the results show that the improvements yielded by the three models are larger in the UAE and Qatar cases than in the case of Saudi Arabia.

This reveals that COVID-19 news has an impact on stock market returns in the UAE, Qatar, Saudi Arabia, and Oman but has no impact on the stock market in Bahrain. It also indicates that COVID-19 news has very little impact on the stock market in Oman compared to that on the markets in the UAE, Qatar, and Saudi Arabia. Furthermore, the impacts on the stock markets in the UAE and Qatar are greater than the impact on the market in Saudi Arabia. This finding coincides with that of the study by Al-Maadid et al. (2020), which found that the leading stock markets (in the UAE, Qatar, and Saudi Arabia) are more sensitive to news in general than the markets in Oman and Bahrain. A possible explanation for this difference maybe be related to the characteristics of these markets. The markets of Bahrain and Oman are the smallest in terms of their total market capitalizations and trading values. Table 8 compares the market capitalizations and trading values of these markets by the end of 2019, i.e., before the COVID-19 pandemic.

It is also worth mentioning that according to the MSCI classification, the markets in Oman and Bahrain are considered frontiers, while the markets in the UAE, Qatar, and Saudi Arabia are considered emerging markets. The markets in Qatar and the UAE were upgraded from frontier to emerging markets in 2014, while the market in Saudi Arabia was upgraded in 2019. Usually, frontier markets have less liquidity than emerging markets. Less liquidity with small market capitalization and small daily trading values, such as in the cases of Oman and Bahrain, allows large institutional investors in the market to control price changes and reduce volatility; this is usually achieved by the institutions that are owned by the government. Furthermore, comparing the individual daily trading value percentages to those of the institutions, the differences are found to be greater in the UAE, Qatar, and Saudi markets than in Oman and Bahrain. Previous studies (Li et al., 2017, Merli and Roger, 2013) found that individual investors herd more than institutional investors during crises. Accordingly, the levels of herding among the investors in the three larger markets (the UAE, Qatar, and Saudi Arabia) are expected to be higher than those in the markets of Bahrain and Oman. Because herding behavior is increased by investors' fears, its impact on stock markets was more visible during COVID-19 due to the uncertainty regarding the future of the global economy.
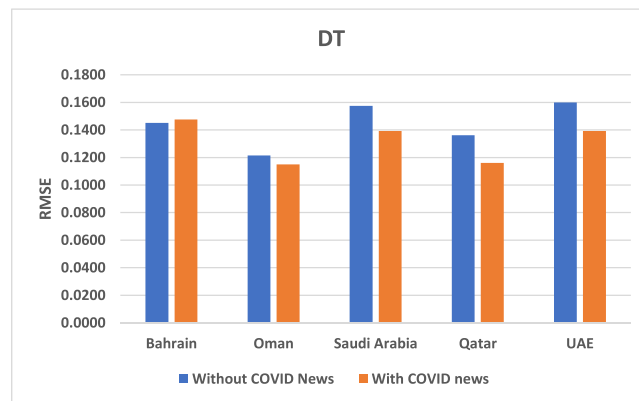
(a) DT
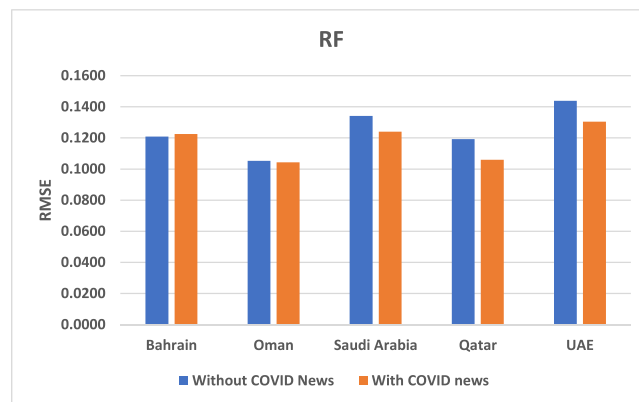
(b) RF
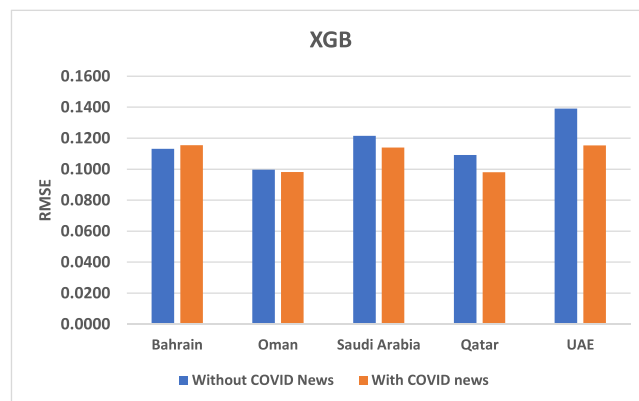
(c) XGB

Fig. 2. Comparison between the RMSE of the test data of the models trained with and without COVID-19-related news features.

Table 8

Comparing GCC stock markets.

Market	Market capitalization (in billion $)	Trading value (in billion $)	Turnover ratio
Saudi Arabia	2398.85	237.6	0.1
Qatar	160.05	18.6	0.12
UAE	246.64	29	0.12
Oman	48.79	1.9	0.04
Bahrain	26.88	0.75	0.03

Table 9
Feature importance for the impacted markets.

Features:	Oman	Saudi Arabia	Qatar	UAE
Business News	0.02558	0.05800	0.04507	0.06262
Google Search Trends	0.13224	0.05392	0.09073	0.13040
Negative Business News	0.00000	0.16186	0.00000	0.00002
Negative Twitter News	0.00963	0.05078	0.04458	0.02068
Number of Cases - Global	0.13113	0.08283	0.03878	0.05702
Number of Cases - Local	0.01332	0.08565	0.07598	0.03904
Number of Deaths - Global	0.04256	0.05295	0.07695	0.04521
Number of Deaths - Local	0.00665	0.01392	0.00254	0.01244
Number of Recoveries - Local	0.02190	0.06900	0.07154	0.09049
Twitter News	0.02010	0.08693	0.34055	0.09392

Additionally, the economies of these countries vary in terms of their degrees of openness and connectedness with the global economy; for example, one of the sectors most influenced during COVID-19 was the real state sector in the UAE because Chinese businessmen are major investors in this sector. Additionally, the cancellation of large events such the Hajj (pilgrimage) in Saudi Arabia and Dubai's 2020 Expo resulted in the loss of hundreds of millions of dollars (Salman and Ali, 2021). Accordingly, the investors in these countries are expected to be more sensitive to any global news that might affect the economy and stock markets.

Table 9 and Fig. 3 illustrate the importance of COVID-19 news features to the impacted markets. The results show that in the case of the UAE, the top three features are Google search trends, Twitter news, and the number of local recoveries. In the case of Qatar, the top three features are Twitter news, Google search trends, and the number of global deaths. In the case of Saudi Arabia, the top three features are negative business news, Twitter news, and the number of local cases. In the case of Oman, Google search trends are the most important, followed by the number of global cases and the number of global deaths.

We note that Twitter news has a significant impact on the stock market in Qatar. Its importance is 34% and represents the top indicator. It also greatly impacts the stock markets in Saudi Arabia (8.6%) and the UAE (9%), and it is ranked second in terms of importance in both cases but is only slightly important in the case of Oman (2%). Google search trends also exhibit high importance in the cases of the UAE (13%), Oman (13%), and Qatar (9%). Negative business news has high importance in the case of Saudi Arabia (16%) but has no importance in the other impacted markets. These differences among the impacted markets in terms of feature importance might be related to the different sources of news that investors pay attention to in these markets, which are affected by a variety of factors, such as internet penetration and social media popularity. According to a report (Mashino, 2019), Qatar and the UAE have the highest rates of social media use among their populations, followed by Saudi Arabia, while Oman has the lowest rate. More research is needed to reveal the differences among the individual investors in these markets.

On the other hand, we note that the number of local death cases has a very low impact on the four markets because the general numbers of coronavirus-related deaths in the GCC countries were lower than those in other countries, such as China, the U.S., and European countries.

In contrast with the study of Ashraf (2020b), who researched the impact of COVID-19 news on the GCC during the period between April 1st, 2020, and June 26th, 2020, our study covers a longer period, so it captures the periods before any local cases were reported and during the global stock decline, as well as the market recovery period. This might explain the inconsistency in the findings, where we find that the number of local deaths has a low impact on the four markets. Furthermore, the study in Ashraf (2020b) did not investigate the impact on each market individually.

6. Conclusions

In this study, we use three machine learning models, namely, DTs, an RF, and XGB, to investigate the impacts of different types of COVID-19-related news on the stock markets in the GCC countries. The results indicate that the stock markets in the UAE, Qatar, Saudi Arabia, and Oman were influenced by news related to COVID-19, but the stock market in Bahrain was not impacted. This finding is consistent across all three models, and the results show that the XGB model has the highest accuracy, followed by the RF, while the DT model has the lowest accuracy. Furthermore, the results reveal that the UAE and Qatar were impacted more than Saudi Arabia, and the impact was nonsignificant in the case of Oman. This difference might be due to the differences among these markets in terms of their sizes, types, and investor types. The markets in Bahrain and Oman are the smallest markets in the GCC countries; they are classified as frontiers, while the markets in the UAE, Qatar and Saudi Arabia are considered emerging markets. Usually, most of the daily trading that takes place in the markets of Oman and Bahrain is achieved by institutional investors who can control the prices and manage the volatility in these small markets. On the other hand, individual investors have the main role in the three largest markets, so hoarding behavior is more pronounced in these markets during crises.

Overall, the results of this study demonstrate that the GCC markets vary in terms of their responses to COVID-19-related news, which has important implications for investors in these markets, who should diversify their portfolios across the markets. These results can be used by investors and portfolio managers to develop risk management strategies that take news related to large global events or disasters into consideration. These strategies include developing monitoring and analysis tools to track the intensity of the global news volume during crises and assess its impact on GCC stock markets. To mitigate risk, different types of news sources should be included, including traditional outlets and social media.
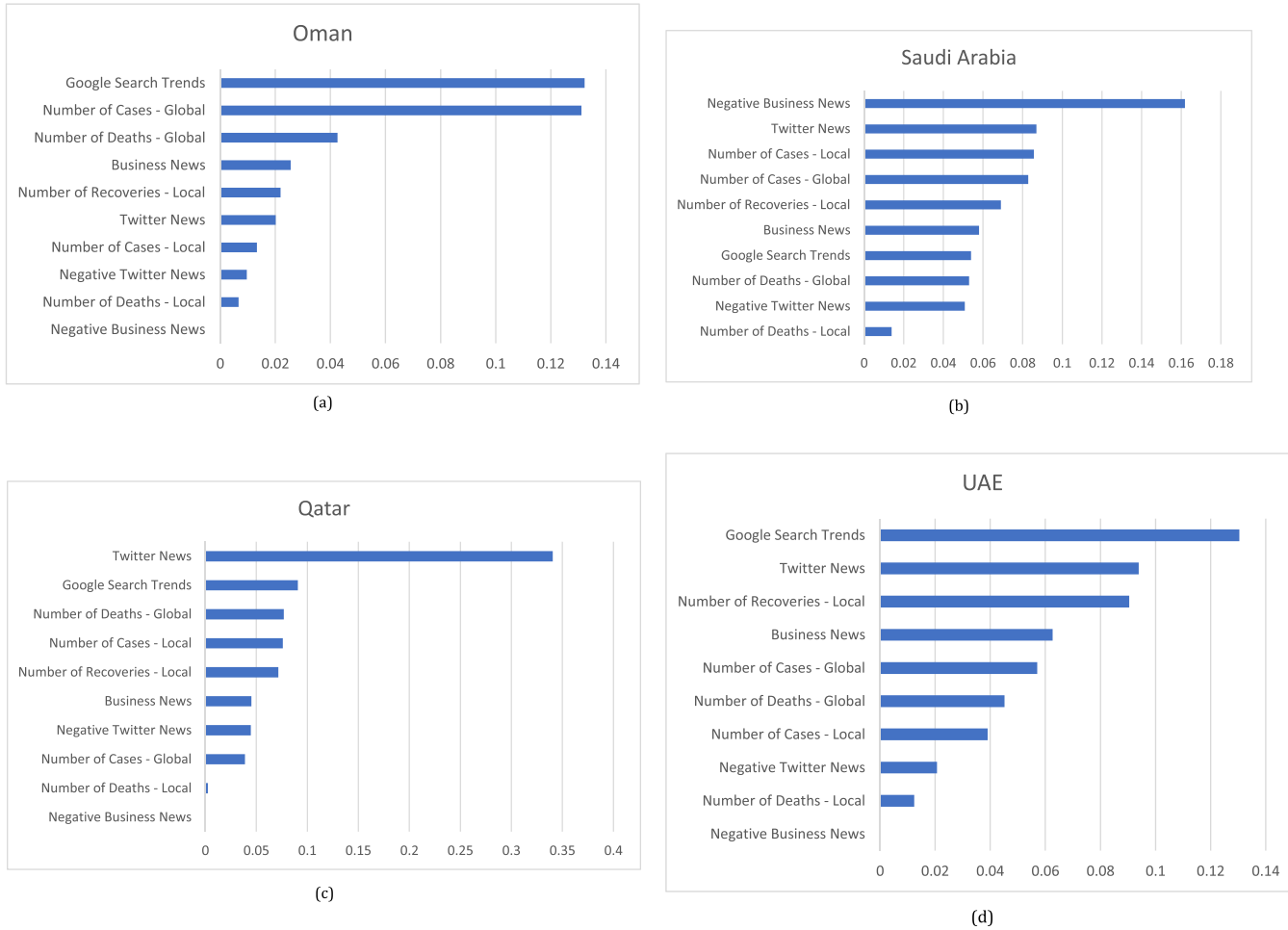
Fig. 3. Feature importance of impacted markets.

Furthermore, the results indicate that the impacted markets were influenced differently by various news types; for example, Twitter news greatly impacts the stock markets in Qatar, the UAE, and Saudi Arabia. However, negative business news has a high impact only on the Saudi Arabian market. Future work can be performed to further investigate the relationships between Twitter news and the stock markets in the GCC over a longer period beyond the COVID-19 pandemic. Additionally, more research is needed to investigate the impacts of COVID-19 news on the different sectors in these markets. One of the limitations in this study is that only U.S. stock volatility (VIX) is considered for capturing the spillover from international markets to the GCC markets; contagion from other markets (such as China) can be addressed in future research.

Acknowledgements

References

Abuzayed, B., Al-Fayoumi, N., 2021. Risk spillover from crude oil prices to GCC stock market returns: new evidence during the COVID-19 outbreak. North Am. J. Econ. Financ. 58, 101476 https://doi.org/10.1016/j.najef.2021.101476.

Adachi, Y., Masuda, M., Takeda, F., 2017. Google search intensity and its relationship to the returns and liquidity of Japanese startup stocks. Pac. Basin Financ. J. 46, 243–257. https://doi.org/10.1016/j.pacfin.2017.09.009.

Akhtaruzzaman, M., Boubaker, S., Sensoy, A., 2020. Financial contagion during COVID–19 crisis. Financ. Res. Lett., 101604 https://doi.org/10.1016/j.frl.2020.101604.

Akhtaruzzaman, M., Boubaker, S., Umar, Z., 2021. COVID–19 media coverage and ESG leader indices. Financ. Res. Lett. 45, 102170 https://doi.org/10.1016/j.frl.2021.102170.

Al-Awadhi, A.M., Alsaifi, K., Al-Awadhi, A., Alhammadi, S., 2020. Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. J. Behav. Exp. Financ. 27, 100326 https://doi.org/10.1016/j.jbef.2020.100326.

Alhazbi, S., Said, A. Ben, & Al-Maadid, A. , 2020. Using deep learning to predict stock movements direction in emerging markets: the case of Qatar stock exchange. In: Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), pp. 440–444. ⟨https://doi.org/10.1109/ICIoT48696.2020.9089616⟩.

Al-Maadid, A., Caporale, G.M., Spagnolo, F., Spagnolo, N., 2020. The impact of business and political news on the GCC stock markets. Res. Int. Bus. Financ. 52, 101102 https://doi.org/10.1016/j.ribaf.2019.101102.

Alqahtani, A., Klein, T., Khalid, A., 2019. The impact of oil price uncertainty on GCC stock markets. Resour. Policy 64, 101526. https://doi.org/10.1016/j.resourpol.2019.101526.

Alqahtani, A., Bouri, E., Vo, X.V., 2020. Predictability of GCC stock returns: the role of geopolitical risk and crude oil returns. Econ. Anal. Policy 68, 239–249. https://doi.org/10.1016/j.eap.2020.09.017.

Ambros, M., Frenkel, M., Huynh, T.L.D., Kilinc, M., 2020. COVID-19 pandemic news and stock market reaction during the onset of the crisis: evidence from high-frequency data. Appl. Econ. Lett. 1–4. https://doi.org/10.1080/13504851.2020.1851643.

Ashraf, B.N., 2020a. Economic impact of government interventions during the COVID-19 pandemic: international evidence from financial markets. J. Behav. Exp. Financ. 27, 100371 https://doi.org/10.1016/j.jbef.2020.100371.

Ashraf, B.N., 2020b. Stock markets' reaction to COVID-19: cases or fatalities? Res. Int. Bus. Financ. 54, 101249 https://doi.org/10.1016/j.ribaf.2020.101249.

Baek, S., Lee, K.Y., 2021. The risk transmission of COVID-19 in the US stock market. Appl. Econ. 53 (17), 1976–1990. https://doi.org/10.1080/00036846.2020.1854668.

Baek, S., Mohanty, S.K., Glambosky, M., 2020. COVID-19 and stock market volatility: an industry level analysis. Financ. Res. Lett. 37, 101748 https://doi.org/10.1016/j.frl.2020.101748.

Bahrini, R., Filfilan, A., 2020. Impact of the novel coronavirus on stock market returns: evidence from GCC countries. Quant. Financ. Econ. 4 (4), 640–652. https://doi.org/10.3934/QFE.2020029.

Beetsma, R., Giuliodori, M., de Jong, F., Widijanto, D., 2013. Spread the news: the impact of news on the European sovereign bond markets during the crisis. J. Int. Money Financ. 34, 83–101. https://doi.org/10.1016/j.jimonfin.2012.11.005.

Benlagha, N., El Omari, S., 2021. What determines the dependence between stock markets – crisis or financial and economic fundamentals? Appl. Econ. 1–19. https://doi.org/10.1080/00036846.2021.1951443.

Bhattacharjee, I., & Bhattacharja, P. , 2019. Stock price prediction: a comparative study between traditional statistical approach and machine learning approach. In: Proceedings of the 2019 4th International Conference on Electrical Information and Communication Technology (EICT), pp. 1–6. ⟨https://doi.org/10.1109/EICT48899.2019.9068850⟩.

Bourdeau-Brien, M., Kryzanowski, L., 2017. The impact of natural disasters on the stock returns and volatilities of local firms. Q. Rev. Econ. Financ. 63, 259–270. https://doi.org/10.1016/j.qref.2016.05.003.

Buckman, S.R., Shapiro, A.H., Sudhof, M., Wilson, D.J., 2020. News sentiment in the time of COVID-19. FRBSF Econ. Lett. 19.

Burggraf, T., Fendel, R., Huynh, T.L.D., 2020. Political news and stock prices: evidence from Trump's trade war. Appl. Econ. Lett. 27 (18), 1485–1488. https://doi.org/10.1080/13504851.2019.1690626.

Cepoi, C.-O., 2020. Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil. Financ. Res. Lett. 36, 101658 https://doi.org/10.1016/j.frl.2020.101658.

Chan, K.F., Gray, P., 2018. Volatility jumps and macroeconomic news announcements. J. Futures Mark. 38 (8), 881–897. https://doi.org/10.1002/fut.21922.

Chan, W.S., 2003. Stock price reaction to news and no-news: drift and reversal after headlines. J. Financ. Econ. 70 (2), 223–260. https://doi.org/10.1016/S0304-405X(03)00146-6.

Chen, T., Guestrin, C. , 2016. XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ⟨https://doi.org/10.1145/2939672.2939785⟩.

Donadelli, M., Kizys, R., Riedel, M., 2017. Dangerous infectious diseases: bad news for main street, good news for wall street? J. Financ. Mark. 35, 84–103. https://doi.org/10.1016/j.finmar.2016.12.003.

Duan, Y., Liu, L., Wang, Z., 2021. COVID-19 sentiment and the Chinese stock market: evidence from the official news media and Sina Weibo. Res. Int. Bus. Financ. 58, 101432 https://doi.org/10.1016/j.ribaf.2021.101432.

Ekinci, C., Bulut, A.E., 2021. Google search and stock returns: a study on BIST 100 stocks. Glob. Financ. J. 47, 100518 https://doi.org/10.1016/j.gfj.2020.100518.

Engelhardt, N., Krause, M., Neukirchen, D., Posch, P., 2020. What drives stocks during the corona-crash? News attention vs. rational expectation. Sustainability 12 (12), 5014. https://doi.org/10.3390/su12125014.

Goodell, J.W., 2020. COVID-19 and finance: agendas for future research. Financ. Res. Lett. 35, 101512 https://doi.org/10.1016/j.frl.2020.101512.

Goodell, J.W., Huynh, T.L.D., 2020. Did Congress trade ahead? Considering the reaction of US industries to COVID-19. Financ. Res. Lett. 36, 101578 https://doi.org/10.1016/j.frl.2020.101578.

Harjoto, M.A., Rossi, F., Lee, R., Sergi, B.S., 2020. How do equity markets react to COVID-19? Evidence from emerging and developed countries. J. Econ. Bus., 105966 https://doi.org/10.1016/j.jeconbus.2020.105966.

Haroon, O., Rizvi, S.A.R., 2020a. Flatten the curve and stock market liquidity – an inquiry into emerging economies. Emerg. Mark. Financ. Trade 56 (10), 2151–2161. https://doi.org/10.1080/1540496X.2020.1784716.

Haroon, O., Rizvi, S.A.R., 2020b. COVID-19: media coverage and financial markets behavior – a sectoral inquiry. J. Behav. Exp. Financ. 27, 100343 https://doi.org/10.1016/j.jbef.2020.100343.

Hung, N.T., 2021. Financial connectedness of GCC emerging stock markets. Eurasia Econ. Rev. https://doi.org/10.1007/s40822-021-00185-2.

Kamal, M.R., Chowdhury, M.A.F., Hosain, M.M., 2021. Stock market reactions of maritime shipping industry in the time of COVID-19 pandemic crisis: an empirical investigation. Marit. Policy Manag. 1–16. https://doi.org/10.1080/03088839.2021.1954255.

Lee, H.S., 2020. Exploring the initial impact of COVID-19 sentiment on US stock market using big data. Sustainability 12 (16), 6648. https://doi.org/10.3390/su12166648.

Li, K., 2018. Reaction to news in the Chinese stock market: a study on Xiong'an new area strategy. J. Behav. Exp. Financ. 19, 36–38. https://doi.org/10.1016/j.jbef.2018.03.004.

Li, Q., Chen, Y., Wang, J., Chen, Y., Chen, H., 2018. Web media and stock markets: a survey and future directions from a big data perspective. IEEE Trans. Knowl. Data Eng. 30 (2), 381–399. https://doi.org/10.1109/TKDE.2017.2763144.

Li, W., Rhee, G., Wang, S.S., 2017. Differences in herding: individual vs. institutional investors. Pac. -Basin Financ. J. 45, 174–185. https://doi.org/10.1016/j.pacfin.2016.11.005.

Loh, W., 2011. Classification and regression trees. WIREs Data Min. Knowl. Discov. 1 (1), 14–23. https://doi.org/10.1002/widm.8.

Mamaysky, H., 2020. Financial markets and news about the coronavirus. SSRN Electron. J. https://doi.org/10.2139/ssrn.3565597.

Mashino, I., 2019. The spread of social media in the GCC and the potential for its utilization in B2C business. ⟨https://www.mitsui.com/mgssi/en/report/detail/__icsFiles/afieldfile/2019/06/25/1903e_mashino_e.pdf⟩.

Mensi, W., Hammoudeh, S., Yoon, S.-M., Balcilar, M., 2017. Impact of macroeconomic factors and country risk ratings on GCC stock markets: evidence from a dynamic panel threshold model with regime switching. Appl. Econ. 49 (13), 1255–1272. https://doi.org/10.1080/00036846.2016.1217305.

Merli, M., Roger, T., 2013. What drives the herding behavior of individual investors? Finance 34 (3), 67. https://doi.org/10.3917/fina.343.0067.

Nguyen, C.P., Schinckus, C., Hong Nguyen, T.V., 2019. Google search and stock returns in emerging markets. Borsa Istanb. Rev. 19 (4), 288–296. https://doi.org/10.1016/j.bir.2019.07.001.

Okorie, D.I., Lin, B., 2020. Stock markets and the COVID-19 fractal contagion effects. Financ. Res. Lett., 101640 https://doi.org/10.1016/j.frl.2020.101640.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Salisu, A.A., Vo, X.V., 2020. Predicting stock returns in the presence of COVID-19 pandemic: the role of health news. Int. Rev. Financ. Anal. 71, 101546 https://doi.org/10.1016/j.irfa.2020.101546.

Salman, A., Ali, Q., 2021. Covid-19 and its impact on the stock market in GCC. J. Sustain. Financ. Invest. 1–17. https://doi.org/10.1080/20430795.2021.1944036.

Sinha, N.R., 2016. Underreaction to news in the US Stock Market. Q. J. Financ. 06 (02), 1650005 https://doi.org/10.1142/S2010139216500051.

Smales, L.A., 2014. Non-scheduled news arrival and high-frequency stock market dynamics. Res. Int. Bus. Financ. 32, 122–138. https://doi.org/10.1016/j.ribaf.2014.03.006.

Stanković, J., Marković, I., Stojanović, M., 2015. Investment strategy optimization using technical analysis and predictive modeling in emerging markets. Procedia Econ. Financ. 19, 51–62. https://doi.org/10.1016/S2212-5671(15)00007-6.

Ulussever, T., Demirer, R., 2017. Investor herds and oil prices evidence in the Gulf cooperation council (GCC) equity markets. Cent. Bank Rev. 17 (3), 77–89. https://doi.org/10.1016/j.cbrev.2017.08.001.

Zaremba, A., Kizys, R., Aharon, D.Y., Demir, E., 2020. Infected markets: novel coronavirus, government interventions, and stock return volatility around the globe. Financ. Res. Lett. 35, 101597 https://doi.org/10.1016/j.frl.2020.101597.

Zhang, Dahai, Qian, L., Mao, B., Huang, C., Huang, B., Si, Y., 2018. A data-driven design for fault detection of wind turbines using random forests and XGboost. IEEE Access 6, 21020–21031. https://doi.org/10.1109/ACCESS.2018.2818678.

Zhang, Dayong, Hu, M., Ji, Q., 2020. Financial markets under the global pandemic of COVID-19. Financ. Res. Lett. 36, 101528 https://doi.org/10.1016/j.frl.2020.101528.