Hindawi
Complexity
Volume 2022, Article ID 4079524, 12 pages
https://doi.org/10.1155/2022/4079524

Research Article

Developing Machine Learning Techniques to Investigate the Impact of Air Quality Indices on Tadawul Exchange Index

Dania AL-Najjar [1], Hazem AL-Najjar [2], Nadia Al-Rousan [3] and Hamzeh F. Assous [1]

[1] *Finance Department, School of Business, King Faisal University, Al Ahsa, Saudi Arabia*
[2] *Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey*
[3] *MIS Department, Faculty of Business, Sohar University, Sohar, Oman*

Correspondence should be addressed to Dania AL-Najjar; dalnajjar@kfu.edu.sa

Received 27 June 2022; Accepted 9 September 2022; Published 6 October 2022

Academic Editor: Gang Jin Wang

The air quality index (AQI) can be described using different pollutant indices. Many investigators study the effect of stock prices and air quality in the field to show if there is a relationship between changing the stock market and the concentration of various pollutants. This study aims to find a relationship between Saudi Tadawul All Share Index (TASI) and multiple pollutants, including particulate matter (PM10), ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), carbon monoxide (CO), and AQI. Based on tree models, the relationship is created using linear regression and two prediction models, Chi-square Automatic Interaction Detection (CHAID), and CR-Tree. In order to achieve the target of this research, the TASI dataset relates to six pollutants using time; afterward, the new dataset is divided into three parts—test, validate, and train—after eliminating the outlier data. In order to test the performance of two prediction models, R^2 and various error functions are used. The linear regression model results found that PM10, NO_2, CO, month, day, and year are significant, whereas O_3, SO_2, and AQI indices are insignificant. The test dataset showed that R^2 scores are 0.995 and 0.986 for CR-Tree and CHAID, respectively, with RMSE values of 387 and 227 for CR-Tree and CHAID, respectively. The prediction models showed that the CHAID model performed better than CR-Tree because it used only three indices, namely, PM10, AQI, and O_3, with year and month. The results indicated an effect between changing TASI and the three pollutants, PM10, AQI, and O_3.

1. Introduction

Climate change is an emergency beyond national borders; it is a worldwide situation where only collaboration between countries can solve it. In Paris, held in December 2015, the Paris climate agreement or COP21, started on 4 November 2016, with 195 countries signing it, and 190 approved it as of January 2021, as mentioned by the United Nations Framework Convention on Climate Change (UNFCCC). This agreement provides a robust pathway leading the global effort to achieve sustainable development goals and shift toward a net-zero emissions world. The primary purpose of this agreement includes promises from all countries to reduce their emissions and work together to adapt to the impacts of climate change. This agreement aims to limit global warming by minimizing greenhouse gas emissions to below 2°C, preferably to 1.5°C, compared to the temperature benchmark set before the beginning of the Industrial Revolution. Besides, the agreement built a structural plan to achieve a balance between atmospheric inputs of greenhouse gases by emission sources and removal into sinks after 2050. The Paris Agreement works on a 5-year cycle of increasingly ambitious climate action carried out by countries. Accordingly, countries in 2020 submitted their plans for climate action, known as nationally determined contributions. Furthermore, this agreement provides an outline for financial, technical, and capacity-building support to those countries that need it. Although the Paris Agreement declared that no new funding targets are required, it is preferable to see developed countries providing funding to help developing countries. This funding mainly supports climate change mitigation and adaptation to climate impacts in less developed countries. This developed–developing support should walk on two parallel sides. On one side, the developed countries support the developing countries through grants, equipment, and technical expertise. On the other side, the less developed countries must enhance their economies and decrease poverty, resulting in solid and direct drops in greenhouse gas emissions.

Saudi Arabia, the world's leading fossil fuel exporter and the world's tenth-largest emitter of carbon dioxide (CO_2), showed an improvement in its pledge to climate action within its borders. By October 2021, the kingdom promised to cut its carbon emissions to net zero by 2060, which means not adding greenhouse gas emissions to the atmosphere. This announcement and many other actions end with upgrading its Climate Action Tracker (CAT) rating from "critically insufficient" to "highly insufficient." Although the Saudi economy depends heavily on oil production and its revenues, the Saudi government endorsed the Circular Carbon Economy (CCE) approach to managing emissions, mitigating climate challenge effects, and making energy systems cleaner and more sustainable. Moreover, the kingdom remains committed to the circular economy's four Rs (i.e., reduce, reuse, recycle, and renew). Saudi Arabia's national emissions have recently decreased after peaking in 2015 due to a decrease in oil consumption in the electricity sector, the COVID-19 impact on the economy, and the pandemic-related global decline in oil demand.

Furthermore, spokespersons of Saudi Arabia announced that the kingdom would achieve the net zero emissions target without affecting the "stability of global energy markets" through many actions. The kingdom aims to use renewable energy to generate 50% of electricity by 2030, as mentioned on the climate change news website. Moreover, Saudi Arabia has announced its aim of planting 450 million trees by 2030, the same as rehabilitating 200 million hectares of degraded land, with collaborative efforts to plant 50 billion trees in the Middle East, as CAT. Finally, the country would use carbon capture—a technology that extracts CO_2 from the air—to help it meet the goal. Several researchers explored the Saudi economy and exchange from different aspects [1, 2].

The air quality index (AQI) measures gas emissions in any country. AQI reports daily air quality and is like the weather; it can change daily or even from hour to hour. It tells to what extent the air is cleaned or polluted and what associated health effects might be a concern for people. The AQI focuses on health effects people may experience within a few hours or days after breathing polluted air. United States Environment Protection Agency (EPA) calculates the AQI for five major air pollutants regulated by the Clean Air Act: ground-level ozone, particle pollution [(particulate matter (PM10)], carbon monoxide (CO), sulfur dioxide (SO_2), and nitrogen dioxide (NO_2). EPA has established national air quality standards for these pollutants to protect public health. Ground-level ozone and airborne particles are the two pollutants that pose the greatest threat to human health in this country. AQI is an index with a scale from 0 to 500. The higher the AQI value, the greater the level of air pollution and the greater the health concern. An AQI value of 50 represents good air quality with little or no potential to affect public health compared to an AQI value of over 300, which means air quality is so unsafe that everyone may have severe effects on their health, as explained on the airnow.gov website. Several scholars focused on investigating the effect of AQI levels on different countries' sectors and financial market indices. In this study, the authors are interested in filling the gap by examining the impact of Saudi AQI on the Saudi stock exchange index, Tadawul All Share Index (TASI).

Based on previous works, the effect of air quality indices on the stock market index is rarely studied by researchers, especially in Saudi Arabia. The studies used various linear and nonlinear models to understand the relationship between the independent variables (air quality indices and date variables) and dependent variables (TASI) without finding the most suitable model. Unfortunately, a few studies have suggested using various variables as an indicator to describe the effect of air pollutants on stock indices. This study presents multiple linear and nonlinear analyses to draw a relationship between independent and dependent variables. Our contribution to this work can be described as follows:

(1) The visualization test and linear regression model are used to show the capability of the linear model in finding a relationship between independent and dependent (TASI) variables
(2) Decision Tree models are used by selecting the two tree models, CR-Tree and Chi-square Automatic Interaction Detection (CHAID), to draw a nonlinear relationship between independent and dependent variables
(3) The prediction model is used to find the most effective pollutants associated with the TASI variable

This study comprises seven sections. Section 2 presents the literature review. Section 3 explains the methodology. Section 4 discusses the analysis. Section 5 describes the conclusions, and Section 6 provides the theoretical and practical implications.

2. Literature Review

This section presents the most related research that links to the environment and AQI, on the one hand, and the Stock Markets Return on the other hand.

2.1. The Environment Pillar Score and Financial Markets Return and Volatility. Many scholars were interested in studying the effect of the environment pillar as one of the sustainability pillars on stock market indices' return and volatility. The findings of Yoo et al. [3] indicate that the environment score significantly positively affects the stock market index, especially for the nonenergy sector. This result contradicts Merckoll and Kvarberg's [4] study, which investigated the effect of the environment pillar score on stock price volatility in Nordic countries from 2010 to 2019. Their results showed a significant negative effect of the environment pillar score on stock price volatility. Besides, Hoepner et al. [5] presented significant adverse effects of a high environment pillar score on the downside risk. Lastly, Eratalay and Angel's [6] study concluded with an insignificant effect of the environment pillar score on the volatility of index return on the sample used.

Nevertheless, other researchers concentrated on other financial indicators, such as the effect of environment pillar score on financial performance and the efficiency of different financial institutes. Furthermore, Alam et al.'s [7] study investigated the effect of the environment pillar score on the efficiency of the Saudi banks and the other three Gulf Cooperation Council (GCC) countries. The outcomes of this study showed a positive *E* score effect on efficiency. Buallay et al. [8] examined the impact of the environment pillar score on Islamic banks' financial performance and found a positive effect of the environment pillar score on banks' performance. Lastly, Yang et al. [9] examined the relationship between financial instability, economic growth, energy consumption, trade openness, urban population, and CO_2 emissions. The findings showed a significant negative impact of financial instability on CO_2 emissions, whereas the increases in economic growth, energy consumption, and urban population are dangerous to the environment.

2.2. Effect of AQI on the Financial Markets. As mentioned before, AQI measures gas emissions in any country. AQI is calculated by considering different air pollutants, namely, SO_2, O_3, nitrogen oxides (NOx), CO_2, PM2.5, and PM10. When the concentration levels of these pollutants in the air increase, they become harmful to all living things around the environment, people, and animals [10, 11]. Many researchers were interested in studying the AQI and the ability of different prediction models to forecast the index using various machine learning models [12–20].

Many researchers have studied the effect of AQI during the COVID-19 period and investigated the positive effect of this pandemic on the worldwide environment, global warming, and gas emissions, as mentioned in Li [21]. Furthermore, Kar et al. [22] indicated that Indian indices had recovered during the pandemic. However, a few scholars studied the relationship between air pollution and the stock market index, as Cunha et al. [23] provided insight into the financial performance of a stock portfolio consisting of carbon-efficient Brazilian companies from 2010 to 2019 through different portfolio metrics. The results indicated that investing in carbon-efficient companies in Brazil contributed positively to the portfolio's performance.

El Ouadghiri et al. [24] studied the effect of public interest in pollution and climate change on weekly returns on indices of sustainability stock in the United States from 2004 to 2018. The findings indicated a significant positive relationship between the public interest in environmental issues and returns on sustainability stock indices in the United States. Moreover, Liu et al. [25] investigated the relationship between investor attention, stock prices, and air pollution. The findings exhibited that air pollution will reduce the polluting companies' stock prices.

According to Ding et al. [26], the results showed a negative relationship between air pollution and company stock returns, as companies in higher air-pollution cities showed lower stock returns. This pollution effect on the return becomes bigger when local investors manage firms with fewer institutional owners and analysts. Jiang et al. [27] investigated the impact of weather and air quality on the equity returns of the Shenzhen Exchange. The findings showed that, based on data from 2005 to 2012, air pollution has significant negative effects on stock returns.

Wu and Guo [28] investigated the relationship between the AQI and the stock yield in key control cities from 2011 to 2016. The findings revealed that severe air pollution significantly negatively influences stock yield. Xu et al. [29] studied the effect of air pollution on stock returns by considering people's awareness of air pollution. The results found that collective awareness of air pollution significantly mediates between air pollution and stock returns. Nguyen and Pham [30] studied the relationship between air pollution and the efficiency of the financial market. The results showed that the stock market anomalies weakened after the severe pollution period.

Other researchers study other essential factors; for example, investors' psychology and political events affect the stock return besides the environment and air pollution. Guo et al. [31] investigated how investor mood and environmental pillars are linked. Their results showed that air pollution negatively and positively affects individual investors buying and selling tendencies. These effects are more remarkable for investors who live in heavily polluted cities and have low investment experience. Antoniuk and Leirvik [32] examined unexpected political events affecting climate-sensitive sectors. These sectors are clean energy, utilities, energy-intensive, fossil energy, and transport. The events either enhance climate change awareness or weaken climate change policy. The results showed that stock market investors quickly adapt to new information regarding climate change.

3. Research Methodology

The research methodology is built after developing different nonlinear prediction models based on neural networks, support vector machines (SVM), Quick Unbiased Efficient Statistical Tree (QUEST), Tree-AS, random forest, CHAID, linear regression, generalized linear regression, and CR-Tree models [33–36]. After extensive study of previous work in the field, the models are considered to extract the most important prediction models in developing stock market prediction models. For brevity, only the highest accurate models are considered to develop the methodology for this research. Initial screening showed that CHAID and CR-Tree are the most accurate models. Meanwhile, linear regression with a visualization test is used to understand the linearity between independent variables (air quality indices and date) and dependent variables (TASI). Finally, this section mainly focuses on data collection and analysis, and the prediction models are designed using two tree models, including CR-Tree and CHAID.

3.1. Data Collection and Analysis. The datasets are collected from two sources to study the relationship between air quality and the Saudi market index (TASI). The stock market index is compiled from the Saudi exchange website and finance.yahoo.com, where air quality indices are collected from two sources: Saudi Arabia General Authority for Meteorology and Environmental Protection and World Air Quality Index Project. The study considered different stations in Riyad, and the six pollutant indices' average values were evaluated. The stations are Khalidiya, Rawabi, Gharbi, Al-Jazeera, and Almurooj. Moreover, only five pollutants indices are measured in Riyad stations, namely, PM10 (μg/m^3), CO (μg/m^3), SO_2 (μg/m^3), NO_2 (μg/m^3), and O_3 (μg/m^3), as well as the AQI, where other pollutants (i.e., PM2.5 and CO_2) are not measured at stations or the collected values are few. The AQI is calculated by finding the maximum value of the five pollutants in an area, as shown in the World Air Quality Index Project. The dataset covers three years, from 2019 and 2021. After collecting the air quality indices and the TASI dataset, the two datasets are combined based on the data with 427 samples. Afterward, the data are cleaned, and the outlier data are removed. The cleaned and nonoutlier dataset is used to develop linear and nonlinear regression models. The descriptive statistic of the cleaned dataset is shown in Table 1. The initial analysis showed that the number of valid cases for pollutants is around 300, where the ranges of pollutants indices are from 1 to 58. The data covered three years with different months and days. In order to analyze the TASI indices, the study used data with the air quality indices to understand the effect of pollutants on the TASI index.

3.2. Develop Prediction Models Based on CR-Tree and CHAID Model. A Pearson correlation analysis is used with a p-value to study the relationship between independent variables (air quality indices) and dependent variables (TASI index). In addition, a visualization test is used to show the relationship between each input variable and the TASI value. Afterward, the linear regression model is adopted upon entering the method. Two tests are used to check the multicollinearity problem, including the VIF and Durbin–Watson tests.

Decision tree models are used to develop a prediction model using data and pollutants. A decision tree is a machine learning model that aims to create a relationship between output and input variables. There are two types of decision trees based on the type of independent variable, namely, categorical and continuous variable decision tree, which considers categorical and continuous variables as the independent variable, respectively. Based on IBM documentation for the SPSS modeler, tree modeling nodes are divided into four types: CR Tree, CHAID, QUEST, and C5.0. Each tree model has its strength and weakness, as discussed in [37]. In order to create a decision tree model, the following assumption must be fit as follows:

Table 1: Descriptive statistics of the cleaned data.

	N	Minimum	Maximum	Mean	Std. deviation
PM10	293	4	58	27.45	10.42
O_3	295	3	52	22.28	10.87
NO_2	292	2	19	9.50	3.41
SO_2	287	1	6	2.23	0.98
CO	275	6	27	14.05	4.91
AQI	300	11	58	30	9.74
TASI	300	6,287	11,512	8,716	1,376
Month	300	1	12	7.31	3.256
Day	300	1	31	15.45	8.750
Year	300	2019	2021	2020.01	0.769

(1) The training dataset is considered a root

(2) The discretized process is applied for continuous variables

(3) Records are distributed recursively

(4) A statistical approach moves from a root node to an internal node

A Sum of Product (SOP) is used to build a decision tree model. Besides, one of the primary challenges in making the decision tree is identifying the independent variable that must be considered a root node for each level. In order to select that, various attributes are considered in the literature, such as entropy, information gain, Gini index, and Chi-square.

In this study, the tree models are used to develop a prediction model. As the tree models are simple to understand and require little data preparation compared to other models, the prediction cost equals the logarithmic of the number of trained data. The decision tree contains many models such as random trees, random forest, C5 tree, Quest tree, Tree-AS, CHAID, and C and R-Tree. In this study, the last two models are selected for many reasons. The CR Tree is considered robust in missing data and large numbers of independent variables; the generated model is straightforward interpretation. On the contrary, the CR tree can generate only binary trees. Therefore, to develop a more robust and accurate model, the CHAID model is used because it can cause nonbinary stress and accept both case weights and frequency variables. In order to build a prediction model, the collected clean dataset is analyzed to check the complexity of the selected data. The dataset is divided into three datasets: train, validate, and test. The data are divided randomly with 70%, 15%, and 15% of the data for training, validating, and testing, respectively, as discussed by Al-Rousan et al. [38], AL-Najjar [39], and AL-Najjar et al. [40]. The training dataset is used to build a prediction model. The developed models are used with validating datasets to improve the prediction rate of the developed models. Afterward, a test dataset is used to check the capability of developed models.

The selected models are evaluated using five metrics, namely, determination coefficient (R^2), error and mean absolute error (MAE), root mean square error (RMSE), mean square error (MSE), mean bias error (MBE), as shown in the following:

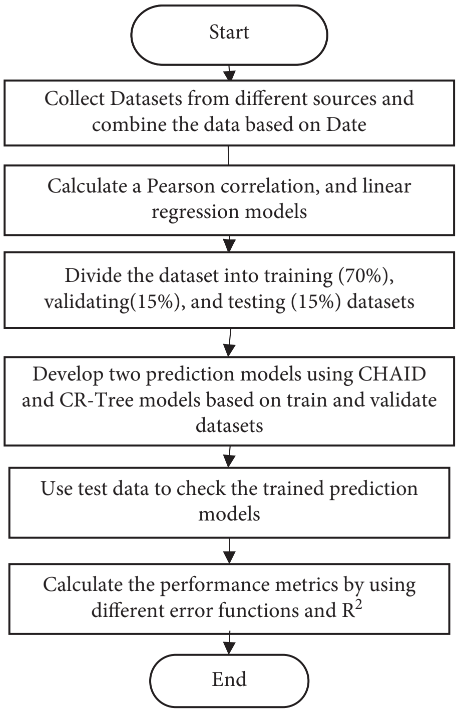

Figure 1: The methodology used to build models.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2},$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=0}^{N} (y_i - \hat{y}_i)^2},$$

$$MSE = \frac{1}{N}\sum_{i=0}^{N} (y_i - \hat{y}_i)^2, \tag{1}$$

$$MAE = \frac{1}{N}\sum_{i=0}^{N} |y_i - \hat{y}_i|,$$

$$MBE = \frac{1}{N}\sum_{i=0}^{N} y_i - \hat{y}_i.$$

where y_i, $\hat{y}_i$, and, $\overline{y}$ are TASI, the predicted TASI, and the mean of TASI, respectively. A flowchart in Figure 1 is considered to summarize all the used methodologies.

4. Result Analysis and Discussion

This section shows various analysis tests to understand the impact of air quality indices on TASI, including correlation analysis, visualization tests, linear regression, and prediction models. Therefore, this section is divided into three subsections: correlation analysis and visualization test, developed linear regression model, and CHAID and CR-Tree prediction models.

4.1. Correlation Analysis and Visualization Test. In order to understand the relationship between air quality indices and TASI, the relationship between air pollutants and TASI is drawn, as shown in Figure 2. The results showed no noticeable relationship between TASI and other pollutants for all the input data. The visual test showed that the lines gradually move with the air quality indices for TASI values from 7,000 to 8,000 and 10,500 to 11,800. Drawing a linear relationship between TASI and air pollutants alone is not possible. Therefore, the date is added to the analysis to support developing linear regression and prediction models.

Before building a linear regression model, a correlation analysis is used to check the possibility of finding a linear regression between independent variables (air quality indices and date) and dependent variables (TASI index). Table 2 shows that the TASI index is insignificantly correlated with PM10, O_3, NO_2, AQI, month, and day, where SO_2 and year are positively correlated with TASI with a correlation coefficient of 0.295 and 0.693, respectively. In addition, CO corrected significantly (correlation level = −0.327, $p < 0.05$) with TASI with an inverse relationship.

On the contrary, one of the crucial tests that must be considered while building a linear regression model is to check the multicollinearity problem; the results in Table 2 reveal that there is a multicollinearity problem, especially between AQI and other pollutants, which may cause unreal linear regression model between the studied variables.

4.2. Developed Linear Regression Model. Linear regression is created to evaluate a linear relationship between air pollutants and TASI, as shown in Tables 3–5. Date and six pollutants indices are used to develop a linear regression model. The results showed that R^2 and standard error are 0.735 and 746, respectively. Moreover, the analysis of variance (ANOVA) test is used to show how TASI changes according to the air quality indices and date level. The results in Table 4 show a significant effect of some pollutant indices and date variables on TASI movement at the $p < 0.05$ level for the three conditions [$F(43828747, 556966) = 79$, $p = 0.000$].

The coefficients of the linear regression are reported in Table 5. The results showed that PM10, NO_2 CO, month, day, and year are significant, whereas O_3, SO_2, and AQI indices are insignificant with TASI. This indicates that the model did not consider the last three indicators in building the model. The preliminary analysis showed which hands were essential to the model but did not show the most critical and most affected variables in the developed linear regression model. In order to check this, unstandardized and standardized coefficients are used. The results showed that the indices of the month, year, and CO had the highest weight on the linear regression model, and the month, year, and PM10 had the highest effect on the linear regression model.

In contrast, to test the multicollinearity problem of the linear regression model, the Durbin–Watson and VIF tests are adopted. The VIF test showed that all the variables did not exceed 10, which indicates that VIF traverses multicollinearity problems, as shown in Table 5. Unfortunately, the Durbin–Watson value is 0.323, meaning a positive multicollinearity problem in the model, as shown in Table 3.

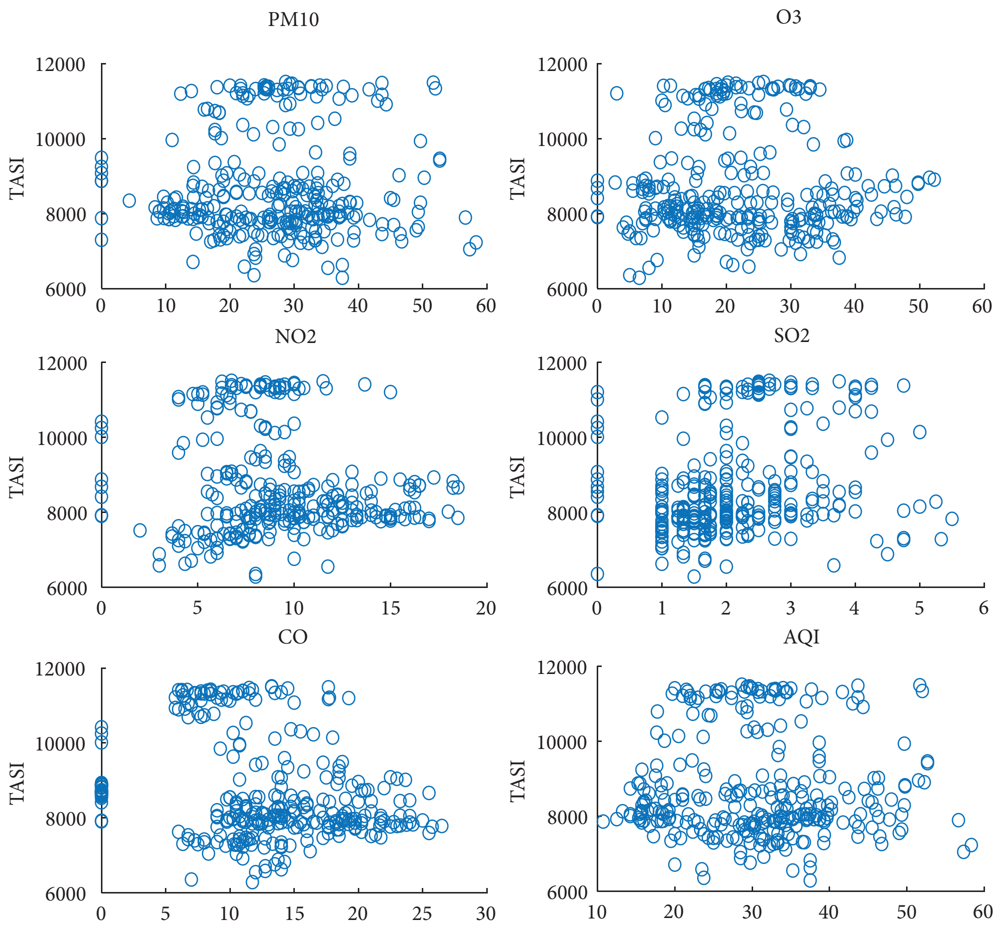

Figure 2: The relationship between the leading pollutant indices and the TASI index.

Table 2: Correlation analysis.

		PM10	O_3	NO_2	SO_2	CO	AQI	Month	Day	Year	TASI
PM10	Corr.	1.000	0.152**	−0.267**	0.163**	−0.176**	0.804**	0.026	0.172**	0.160**	0.043
	Sig.		0.010	0.000	0.006	0.004	0.000	0.660	0.003	0.006	0.466
O_3	Corr.	0.152**	1.000	−0.227**	0.199**	0.194**	0.563**	0.224**	0.005	−0.248**	0.003
	Sig.	0.010		0.000	0.001	0.001	0.000	0.000	0.933	0.000	0.962
NO_2	Corr.	−0.267**	−0.227**	1.000	−.066	0.335**	−0.267**	0.284**	−0.028	−0.320**	−0.113
	Sig.	0.000	0.000		0.265	0.000	0.000	0.000	0.633	0.000	0.053
SO_2	Corr.	0.163**	0.199**	−0.066	1.000	−0.125*	0.165**	0.088	0.048	0.266**	0.295**
	Sig.	0.006	0.001	0.265		0.041	0.005	0.135	0.417	0.000	0.000
CO	Corr.	−0.176**	0.194**	0.335**	−0.125*	1.000	0.002	0.054	−0.035	−0.585**	−0.327**
	Sig.	0.004	0.001	0.000	0.041		0.977	0.370	0.561	0.000	0.000
AQI	Corr.	0.804**	0.563**	−0.267**	0.165**	0.002	1.000	0.085	0.158**	−0.057	−0.013
	Sig.	0.000	0.000	0.000	0.005	0.977		0.140	0.006	0.325	0.823
Month	Corr.	0.026	0.224**	0.284**	0.088	0.054	0.085	1.000	−0.048	−0.422**	0.060
	Sig.	0.660	0.000	0.000	0.135	0.370	0.140		0.409	0.000	0.298
Day	Corr.	0.172**	0.005	−0.028	0.048	−0.035	0.158**	−0.048	1.000	0.027	0.039
	Sig.	0.003	0.933	0.633	0.417	0.561	0.006	0.409		0.639	0.501
Year	Corr.	0.160**	−0.248**	−0.320**	0.266**	−0.585**	−0.057	−0.422**	0.027	1.000	0.693**
	Sig.	0.006	0.000	0.000	0.000	0.000	0.325	0.000	0.639		0.000
TASI	Corr.	0.043	0.003	−0.113	0.295**	−0.327**	−0.013	0.060	0.039	0.693**	1.000
	Sig.	0.466	0.962	0.053	0.000	0.000	0.823	0.298	0.501	0.000	

**Significant at the 0.01 level (two-tailed). *Significant at the 0.05 level (two-tailed).

Table 3: Linear regression summary.

Model	R	R^2	Adjusted R^2	Std. error	Durbin–Watson
1	0.857[a]	0.735	0.726	746	0.323

[a]Predictors: (constant), year, day, AQI, SO_2, month, NO_2, O_3, CO, PM10.

Table 4: ANOVA test for the linear regression model.

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	394,458,723	9	43,828,747	79	0.000[b]
	Residual	142,026,398	255	556,966		
	Total	536,485,121	264			

b:this model is significant at $P<0.01$.

Table 5: Coefficients of the linear regression.

Model		Unstandardized Coefficients		Standardized coefficients			Collinearity Statistics	
		B	std. error	Beta	t	Sig.	tolerance	VIF
1	(Constant)	−4244019	189,007		−22.45	0.00		
	PM10	−30.87	9.62	−0.23	−3.21	0.00	0.20	4.96
	O_3	9.05	7.17	0.07	1.26	0.21	0.36	2.74
	NO_2	43.42	18.82	0.09	2.31	0.02	0.62	1.61
	SO_2	−62.05	52.78	−0.04	−1.18	0.24	0.79	1.27
	CO	68.20	12.57	0.24	5.42	0.00	0.55	1.81
	AQI	21.38	12.23	0.15	1.75	0.08	0.15	6.65
	Month	211.95	19.29	0.45	10.99	0.00	0.62	1.62
	Day	11.09	5.37	0.07	2.06	0.04	0.95	1.05
	Year	2103.86	93.49	1.13	22.50	0.00	0.41	2.43

Table 6: CHAID and CR-Tree prediction results analysis.

	CR-Tree			CHAID		
	Train	Test	Validate	Train	Test	Validate
R^2	0.995	0.972	0.962	0.979	0.986	0.994
MSE	20,240	104,592	149,478	79,746	51,581	21,717
MAE	103	228	255	182	166	112
MBE	0.000	−63.707	41.184	18.177	4.452	21.737
RMSE	142	323	387	282	227	147

Therefore, the linear regression model failed to develop a relationship between air pollutant indices with date variables and TASI values.

4.3. Developed CHAID and CR-Tree Prediction Models. In order to solve the problem in the linear regression model problem and build a system that can predict TASI values accurately and study the most affected pollutants from TASI changes, this section aims to develop two TASI predictions and tree models. Two tree models (i.e., CHAID and CR-Tree) and three datasets (including train, test, and validate) are used. The train results of CR-Tree are 0.995, 20,240, 103, 0.000, and 142, whereas, for the test, the results are 0.962, 149,478, 255, 41.184, and 387 for R^2, MSE, MAE, MBE, and RMSE, respectively, as shown in Table 6. Moreover, the CHAID model results of the training dataset are 0.979, 79,746, 182, 18, and 282, whereas, for the test dataset, the results are 0.986, 51,581, 166, 4, and 227 for R^2, MSE, MAE, MBE, and RMSE, respectively, as shown in Table 6. The results indicated that the CHAID model is more stable and accurate than the CR-Tree model as the prediction rate for testing is higher than the CR-Tree model and the error values are less for all cases.

A visual test is used with the collected data to support the claim of the collected results, as shown in Figures 3 and 4. Figure 3 shows the relationship between the collected data and the TASI values. Three lines are drawn: actual TASI values, CHAID predicted values, and CR-Tree predicted values. The results showed that CHAID and CR-Tree predictors move with TASI values, with CHAID having a more remarkable ability to track increases and decreases in the actual TASI values. In addition, Figure 4 shows the error for each sample using CHAID and CR-Tree predictors; the CHAID results were more accurate than the CR-Tree results.

Moreover, CR-Tree and CHAID used different variables to develop TASI prediction models, as shown in Figure 5. The CR-Tree model used eight variables, including month, day, NO_2, PM10, O_3, AQI, SO_2, and CO, whereas CHAID used only five variables: month, year,

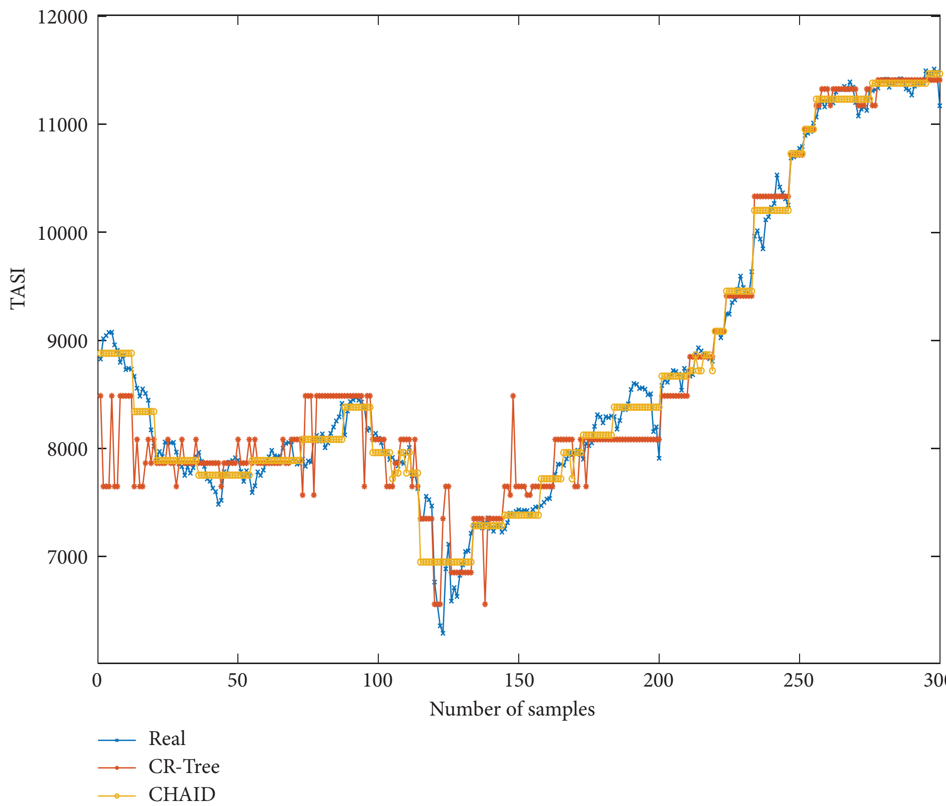

Figure 3: The relationship between all samples and the TASI value using real, CR-Tree, and CHAID models.

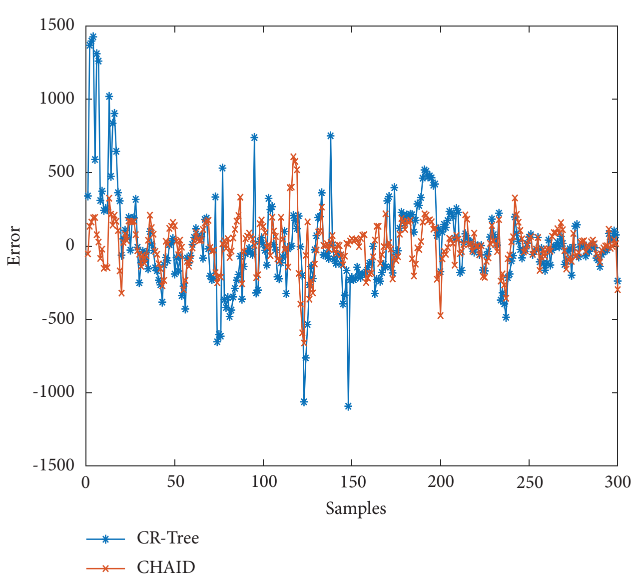

Figure 4: Train error function, testing, and validating samples using CR-Tree and CHAID.

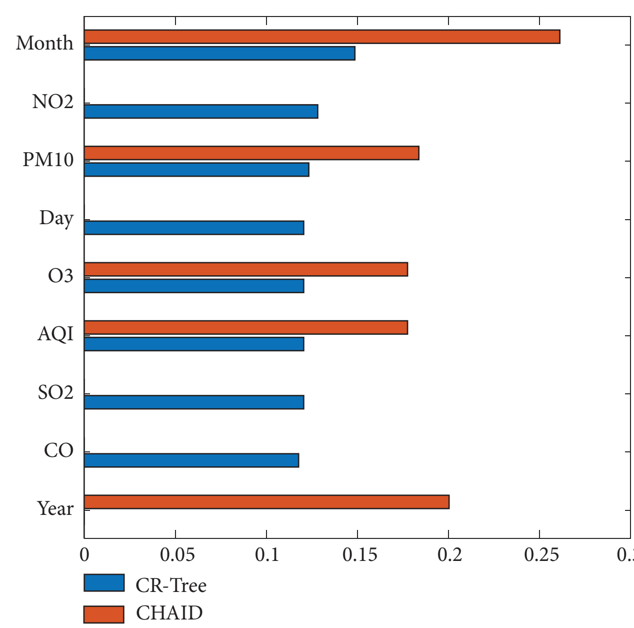

Figure 5: The importance variable for CHAID and CR-Tree models.

PM10, O_3, and AQI. CR-Tree showed that the top four important variables (arranged from most important to least important) are month, NO_2, PM10, and day, whereas the top four important variables in CHAID are month, year, PM10, O_3, and AQI. By combining the collected results and the linear regression analysis, the results found that the CHAID model is more realistic than CR-Tree, as a month, year, and three pollutants indices were included in the developed model.

Furthermore, to understand the complexity of the developed CR-Tree and CHAID prediction models, partial classification trees are explained in Figures 6 and 7, respectively. Tree's depths for CR-Tree and CHAID models are 5 and 3, respectively. CR-Tree model starts by merging 2020

Figure 6: Partial classification tree for TASI CR-Tree predictor.

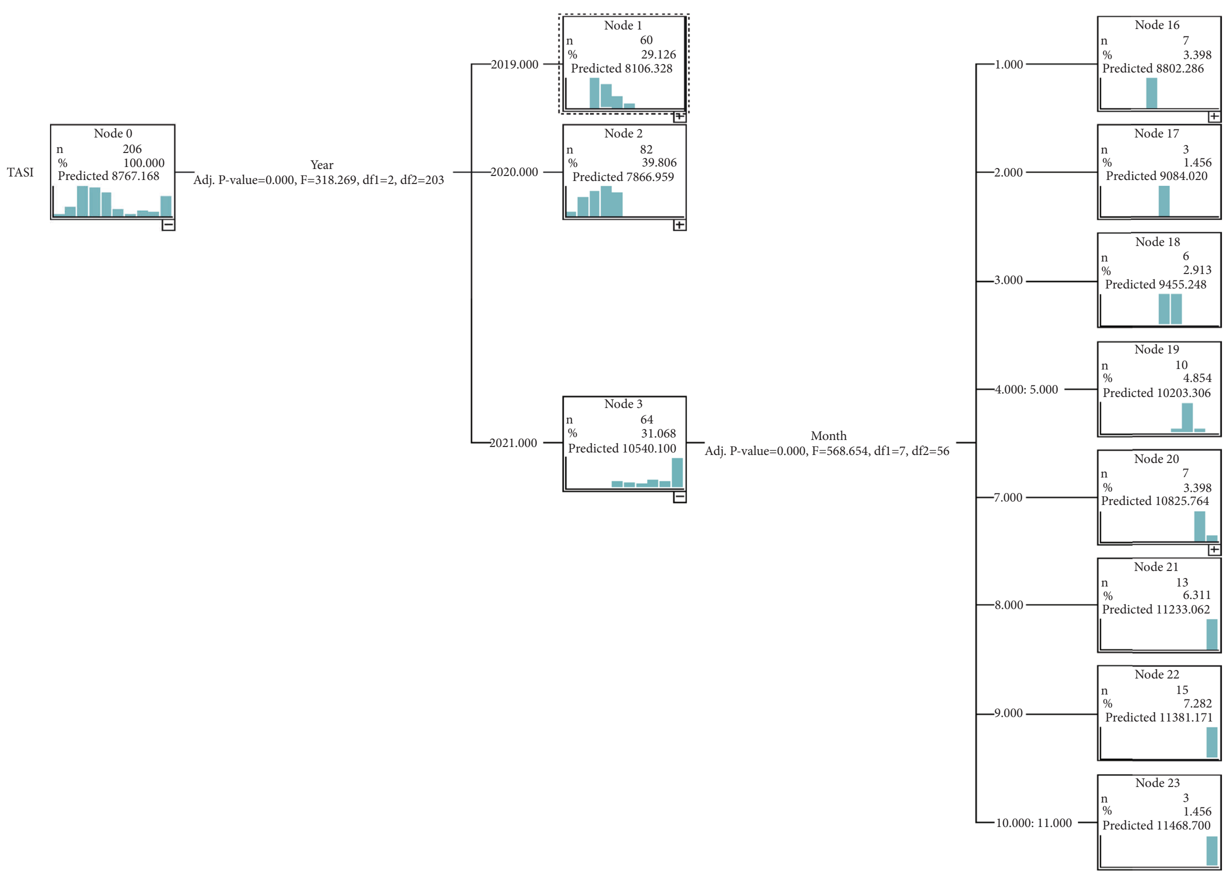

Figure 7: Partial classification tree for TASI CHAID predictor.

and 2019. Afterward, the month variable is used as the main variable for classification. The day, NO_2, and month are used for the second level, whereas the month and CO variables are used for the third and fourth levels, respectively. On the contrary, the CHAID model starts by dividing the data into three parts based on the year variable; then, the month is used to classify the data as the main parameter. Afterward, PM10, AQI, and O_3 are used as the last level for TASI classification.

There is no known research article for authors who predicted Tadawul All Share Index (TASI) using air quality indices as independent variables. Therefore, the researchers compared the best results with most research that predicted the stock market using neural network models described in [36] and [41]. Assous and Al-Najjar [36] reported that the neural network model is the best in stock market predictions. The R^2 and RMSE error values for TASI prediction are 0.96 and 194, respectively. In addition, Malibari et al. [41] predicted the TASI closing index with an accuracy of about 97% and low error values for different batches. The overall results showed that the CHAID model can predict stock markets using date and air quality indices.

To sum up, analytical analysis, visual test, and linear regression model with tree design showed that developing a prediction model using tree models is efficient and accurate in predicting the TASI values. The tests indicated that the CHAID model is correct, robust, stable, realistic, and less complicated than the CR-Tree model. The results revealed a strong relationship between TASI changes and three air quality indices, including AQI, PM10, and O_3.

5. Conclusion and Future Work

The research aims to study the relationship between air pollutant indices with the date and the TASI index (Saudi stock index) in the last three years. The investigation started by collecting air quality indices and TASI values; both datasets are combined based on date. The newly collected dataset is preprocessed to remove the outlier data. As a first step to understanding the relationship between independent variables (air pollutants and date) and dependent variables (TASI value), a linear regression model is used with a visualization test. The results showed a weak linear relationship between air quality indices with date and TASI. Therefore, nonlinear tree models are used to find the relationship between independent and dependent variables. The study used two tree models, including CHAID and CR-Tree. The train results of CR-Tree are 0.995, 20,240, 103, 0.000, and 142, whereas, for the test, the results are 0.962, 149,478, 255, 41.184, and 387 for R^2, MSE, MAE, MBE, and RMSE, respectively, where the CHAID model results of the training dataset are 0.979, 79,746, 182, 18, and 282, whereas, for the test dataset, the results are 0.986, 51581, 166, 4, and 227 for R^2, MSE, MAE, MBE, and RMSE, respectively. The results found that the CHAID model performed better than CR-Tree. The results indicated a nonlinear effect between TASI and three air quality indices, including AQI, PM10, and O_3. CHAID results showed that air quality indices increased the performance of the stock market prediction model, in contrast to [36] and [41], which showed that the neural network model is more accurate than other models in developing stock market prediction models. This gives strong evidence that the changing air quality indices can reflect changes in TASI values. Unfortunately, this will not provide complete evidence about the effect of pollutants' indices on TASI and the direction of the impact, as the decision tree models are nonlinear prediction models. Therefore, extra analysis is required to understand the direction and the important variables that can improve the performance of the prediction models. In future work, the authors aim to improve the prediction model using different machine learning models, including clustering and reinforcement learning. In addition, extra variables such as humidity and temperature are needed to understand their performance on the TASI variable [42].

6. Theoretical and Practical Implications

This study is necessary for academics, scholars, and investors in the field of stock market trading in Saudi Arabia. This study investigates the effect of air pollution with air quality indices on the stock market in the last three years. In order to achieve the target of this study, two tree models are used with different independent variables. Besides, the study only dealt with the general stock index in the Kingdom of Saudi Arabia, which is the Saudi Stock Exchange index (TASI).

Traffic and weather are the main reasons for changing stock trading in some countries. Different pollutants from cars and changing the weather are considered to study, including PM10, O_3, NO_2, SO_2, and CO. In this study, the authors proved that the stock market responds powerfully to the variations in air quality indices in Riyad (the capital of Saudi Arabia). In addition, the results show that climate changes in the atmosphere could cause indirect changes in the price of TASI values.

In addition, policymakers can use this research result to monitor the traffic of cars, power plants, industrial boilers, power generators, pollen, and other sources of O_3 and PM10 pollutants. Furthermore, the results may trigger the shareholders to find suitable dates for selling and buying their shares. Besides, more regulations must be considered on the traffic to decrease the pollutants in the air.

Finally, the societal benefit of the findings of this research can be summarized as follows: (1) the researchers demonstrated a strong influence of PM10, O_3, and AQI on stock market predictions, and this would give an initial indication that pollution could have an initial alarm on reducing or increasing stock trading. (2) Using a CHAID with the date and air pollutants to predict the stock market can improve the accuracy of the stock prediction. Therefore, new shareholders and investors can use this feature to buy and sell shares of stock.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request by Dr. Dania Al-Najjar.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

Acknowledgments

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. 1597).

References

[1] D. Al-Najjar and H. F. Assous, "Key determinants of deposits volume using CAMEL rating system: the case of Saudi banks," *PLoS One*, vol. 16, no. 12, Article ID e0261184, 2021.

[2] H. F. Assous, N. Al-Rousan, D. Al-Najjar, and H. Al-Najjar, "Can international market indices estimate TASI's movements? The ARIMA model," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 6, no. 2, pp. 27–18, 2020.

[3] S. Yoo, A. R. Keeley, and S. Managi, "Does sustainability activities performance matter during financial crises? Investigating the case of COVID-19," *Energy Policy*, vol. 155, Article ID 112330, 2021.

[4] H. Merckoll and O. Kvarberg, "Har ESG-score en Effekt på Aksjeprisvolatilitet i det Nordiske Markedet? En Empirisk Studie," Master's thesis, University of stavanger, Norway, 2021.

[5] A. G. F. Hoepner, I. Oikonomou, Z. Sautner, L. T. Starks, and X. Zhou, "ESG shareholder engagement and downside risk," 2019, https://ssrn.com/abstract=2874252.

[6] M. H. Eratalay and A. P. Cortés Ángel, "The impact of ESG ratings on the systemic risk of European blue-chip firms," *Journal of Risk and Financial Management*, vol. 15, no. 4, p. 153, 2022.

[7] A. W. Alam, H. Banna, and M. K. Hassan, "ESG activities and bank efficiency: are Islamic banks better?" Available at SSRN 4030145, 2021.

[8] A. Buallay, "Is sustainability reporting (ESG) associated with performance? Evidence from the European banking sector," *Management of Environmental Quality: An International Journal*, vol. 30, no. 1, pp. 98–115, 2019.

[9] B. Yang, M. Ali, M. R. Nazir, W. Ullah, and M. Qayyum, "Financial instability and CO2 emissions: cross-country evidence," *Air Quality, Atmosphere & Health*, vol. 13, no. 4, pp. 459–468, 2020.

[10] L. N. Araujo, J. T. Belotti, T. A. Alves, Y. d. S. Tadano, and H. Siqueira, "Ensemble method based on Artificial Neural Networks to estimate air pollution health risks," *Environmental Modelling & Software*, vol. 123, Article ID 104567, 2020.

[11] C. Shi, I. C. Nduka, Y. Yang et al., "Characteristics and meteorological mechanisms of transboundary air pollution in a persistent heavy PM2. 5 pollution episode in Central-East China," *Atmospheric Environment*, vol. 223, Article ID 117239, 2020.

[12] S. Hansun, A. Wicaksana, and M. B. Kristanda, "Prediction of Jakarta City air quality index: modified double exponential smoothing approaches," *Int J Innov Comput Inf Control*, vol. 17, no. 4, pp. 1363–1371, 2021.

[13] L. Hota, P. K. Dash, K. S. Sahoo, and A. H. Gandomi, "Air quality index analysis of Indian cities during COVID-19 using machine learning models: a comparative study," in *Proceedings of the 2021 8th International Conference on Soft Computing & Machine Intelligence (ISCMI)*, pp. 27–31, IEEE, Cario, Egypt, November 2021.

[14] Y. Mo, Q. Li, H. Karimian et al., "A novel framework for daily forecasting of ozone mass concentrations based on cycle reservoir with regular jumps neural networks," *Atmospheric Environment*, vol. 220, Article ID 117072, 2020.

[15] M. Yu, X. Cai, Y. Song, and X. Wang, "A fast-forecasting method for PM2. 5 concentrations based on footprint modeling and emission optimization," *Atmospheric Environment*, vol. 219, Article ID 117013, 2019.

[16] Z. Wang, L. Chen, Z. Ding, and H. Chen, "An enhanced interval PM2. 5 concentration forecasting model based on BEMD and MLPI with influencing factors," *Atmospheric Environment*, vol. 223, Article ID 117200, 2019.

[17] H. Liu and C. Chen, "Prediction of outdoor PM2. 5 concentrations based on a three-stage hybrid neural network model," *Atmospheric Pollution Research*, vol. 11, no. 3, pp. 469–481, 2020.

[18] H. d. He and Wz Lu, "Comparison of three prediction strategies within PM2. 5 and PM10 monitoring networks," *Atmospheric Pollution Research*, vol. 11, no. 3, pp. 590–597, 2020.

[19] S. Zhu, X. Qiu, Y. Yin et al., "Two-step-hybrid model based on data pre-processing and intelligent optimization algorithms (CS and GWO) for NO2 and SO2 forecasting," *Atmospheric Pollution Research*, vol. 10, no. 4, pp. 1326–1335, 2019.

[20] A. Alimissis, K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi, "Spatial estimation of urban air pollution with the use of artificial neural network models," *Atmospheric Environment*, vol. 191, pp. 205–213, 2018.

[21] X. Li, "On the multifractal analysis of air quality index time series before and during COVID-19 partial lockdown: a case study of Shanghai, China," *Physica A: Statistical Mechanics and its Applications*, vol. 565, Article ID 125551, 2021.

[22] S. Kar, I. Ghosh, S. Show et al., "Impact of coronavirus (COVID-19) outbreak on society, air quality, and economy in India: a study of three "P" s of sustainability in India," *Sustainability*, vol. 13, no. 5, p. 2873, 2021.

[23] F. A. F. d. S. Cunha, E. Meira, R. J. Orsato, M. C. Klotzle, and A. F. Lucena, "Do low-carbon investments in emerging economies pay off? Evidence from the Brazilian stock market," *International Review of Financial Analysis*, vol. 74, Article ID 101700, 2021.

[24] I. El Ouadghiri, K. Guesmi, J. Peillex, and A. Ziegler, "Public attention to environmental issues and stock market returns," *Ecological Economics*, vol. 180, Article ID 106836, 2021.

[25] F. Liu, Y. Kang, K. Guo, and X. Sun, "The relationship between air pollution, investor attention and stock prices: evidence from new energy and polluting sectors," *Energy Policy*, vol. 156, Article ID 112430, 2021.

[26] X. Ding, M. Guo, and T. Yang, "Air pollution, local bias, and stock returns," *Finance Research Letters*, vol. 39, Article ID 101576, 2021.

[27] Z. Jiang, R. Gupta, S. Subramaniam, and S. M. Yoon, "The effect of air quality and weather on the Chinese stock: evidence from shenzhen stock exchange," *Sustainability*, vol. 13, no. 5, p. 2931, 2021.

[28] X. Wu, S. Chen, J. Guo, and G. Gao, "Effect of air pollution on the stock yield of heavy pollution enterprises in China's key control cities," *Journal of Cleaner Production*, vol. 170, pp. 399–406, 2018.

[29] M. Xu, Y. Wang, and Y. Tu, "Uncovering the invisible effect of air pollution on stock returns: a moderation and mediation analysis," *Finance Research Letters*, vol. 39, Article ID 101646, 2021.
[30] H. T. Nguyen and M. H. Pham, "Air pollution and behavioral biases: evidence from stock market anomalies," *Journal of Behavioral and Experimental Finance*, vol. 29, Article ID 100441, 2021.
[31] M. Guo, M. Wei, and L. Huang, "Does air pollution influence investor trading behavior? Evidence from China," *Emerging Markets Review*, vol. 50, Article ID 100822, 2022.
[32] Y. Antoniuk and T. Leirvik, "Climate change events and stock market returns," *Journal of Sustainable Finance & Investment*, vol. 26, 2021.
[33] D. Al-Najjar, "The Co-movement between international and emerging stock markets using ANN and stepwise models: evidence from selected indices," *Complexity*, vol. 2022, Article ID 7103553, 14 pages, 2022a.
[34] H. F. Assous, "Prediction of banks efficiency using feature selection method: comparison between selected machine learning models," *Complexity*, vol. 2022, Article ID 3374489, 2022.
[35] H. Al-Najjar, N. Al-Rousan, D. Al-Najjar, H. F. Assous, and D. Al-Najjar, "Impact of COVID-19 pandemic virus on G8 countries' financial indices based on artificial neural network," *Journal of Chinese Economics and Foreign Trade Studies*, vol. 14, no. 1, pp. 89–103, 2021.
[36] H. F. Assous and D. Al-Najjar, "Consequences of COVID-19 on banking sector index: artificial neural network model," *International Journal of Financial Studies*, vol. 9, no. 4, p. 67, 2021.
[37] A. Melsayad and H. A Elsalamony, "Diagnosis of breast cancer using decision tree models and SVM," *International Journal of Computer Application*, vol. 83, no. 5, pp. 19–29, 2013.
[38] N. Al-Rousan, N. A. Mat Isa, M. K. Mat Desa, and H. Al-Najjar, "Integration of logistic regression and multilayer perceptron for intelligent single and dual axis solar tracking systems," *International Journal of Intelligent Systems*, vol. 36, no. 10, pp. 5605–5669, 2021.
[39] D. Al-Najjar, "Impact of the twin pandemics: COVID-19 and oil crash on Saudi exchange index," *PLoS One*, vol. 17, no. 5, Article ID e0268733, 2022b.
[40] D. Al-Najjar, H. F. Assous, H. Al-Najjar, and N. Al-Rousan, "Ramadan effect and indices movement estimation: a case study from eight Arab countries," *Journal of Islamic Marketing*, vol. 7, 2022.
[41] N. Malibari, I. Katib, and R. Mehmood, "Predicting stock closing prices in emerging markets with transformer neural networks: the Saudi stock exchange case," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 12, 2021.
[42] H. Al-Najjar and N. Al-Rousan, "A classifier prediction model to predict the status of Coronavirus CoVID-19 patients in South Korea," *European Review for Medical and Pharmacological Sciences*, vol. 24, 2020.