

*Research article*

## **Analysis Factors Affecting Egyptian Inflation Based on Machine Learning Algorithms**

**Mohamed F. Abd El-Aal\***

Economics department, Faculty of Commerce, Arish University, North Sinai, Egypt

\* **Correspondence:** Email: [Mohammed.fawzy@comm.aru.edu.eg](mailto:Mohammed.fawzy@comm.aru.edu.eg).

**Abstract:** Given that inflation is one of the most important problems facing the Egyptian economy, determining the factors affecting it is very important. Thus, we aim to use machine learning algorithms like Support vector machine (SVM), K-nearest neighbor (KNN), Random Forest (R.F.), Neural Network (ANN), Gradient boosting (G.B.) and decision tree (D.T.) to determine the accurate algorithm and analyze the factors affecting on Egypt inflation. The study found that the G.B. algorithm is the most accurate among the used algorithms and showed that The major significant variables determining inflation in Egypt are the exchange rate (30.5%), gross fixed formation (24.5%) and government expenditure (12.3%). We also found a positive relationship between the inflation rate and government expenditure, money supply, gross domestic product (GDP) growth, gross fixed formation, foreign direct investment, GDP per capita and exchange rate. Furthermore, there is a negative relationship between the inflation rate, household expenditure and the external trade balance.

**Keywords:** inflation rate forecasting, machine learning, neural network, gradient boosting, decision tree

**JEL Codes:** C80, C81, C87

---

**Abbreviations:** GDP: Gross domestic product; ML: Machine learning algorithms; G.B.: Gradient boosting algorithms; SVM: Support vector machine; D.T: Decision tree; KNN: K-nearest neighbor; RNN: Recurrent neural networks; RMSE: Root mean square error; MAE: Mean absolute error; MAPE: Mean absolute percentage error;  $R^2$ : The coefficient of determination

## 1. Introduction

Forecasting inflation is an extremely difficult task for the decision-making processes of economic players. The direction of the overall price level would affect the future investment decisions of corporations. Thus, the dividend yield of the investments depends greatly on inflation. In addition to influencing households' decisions on labor supply, the inflation rate also plays a crucial role. Due to the effect of inflation on their real incomes, they may adjust their choices. Each participant in the market would change their assessments based on the expectation of inflation (Monteforte & Moretti, 2013).

Since the early 1990s, the inflation rate targeting approach was implemented. Central banks have announced a target inflation rate for a given period and used various instruments to obtain the target inflation level (Iversen et al., 2016). Moreover, while monetary policy adjustments are susceptible to lags, central banks must have accurate inflation projections to determine their desired inflation rates. Thus, the precision of the inflation rate estimates a considerable challenge for monetary authorities (Baybuza, 2018). The constant inflation change in emerging nations discourages long-term investments and reduces economic agents' assets and liability matures (Garcia et al., 2017).

The most challenging component of inflation prediction is determining which variables are the most predictive. Consequently, even though it is crucial to comprehend the relevant signs for inflation prediction in any economy, it is more difficult to examine inflation dynamics in emerging economies. Utilized analytical models face the issue of selecting the most effective methods (Groen et al., 2013).

Egypt, like more developing countries, suffers from the problem of inflation; during the period 1976 to 2022, the inflation rate reached its highest level of 29.5 % in 2017, and it is possible that this will exceed soon after reducing the value of the pound against the dollar in the year 2022. Because of the problems that countries face due to inflation, the importance of this study appeared to analyze its causes and determine the most important variables that affect it to prepare an appropriate economic policy to confront this problem.

In the big data environment, developing machine learning techniques (ML) overcomes the overfitting problem, leading to substantial advances in data analysis (Baybuza, 2018). We use machine learning algorithms like SVM, KNN, R.F., ANN, G.B. and D.T. to obtain accurate results. After the parallel algorithm in prediction is determined, the relationship between the inflation rate and explanatory variables is determined. On the other hand, the recent introduction of machine learning techniques enables us to execute strong modeling techniques in inflation forecasting operations. These models use various macroeconomic variables and diverse financial measurements to examine inflation fluctuations (Yadav et al., 2019).

## 2. Literature review

Friedman's (1968) findings indicate that inflation is a worldwide monetary issue. Controlling the issue and stabilizing the economy for quick economic expansion and growth is the top goal of every nation. Due to its importance to economic entities and policymakers, inflation forecasting receives increasing attention and is the issue of a growing number of research articles. Many studies have examined various techniques for forecasting inflation in both established and developing nations. Although it is feasible to divide the enormous literature into numerous subcategories, we summarize the methodological parameters of extant studies. Briefly, references to prior research are made in one portion of the present study.

This study (Yolanda, 2017) investigates the connection between inflation and interest rates, foreign currency rates, oil and gold prices and money supply. The outcomes of this study reveal that interest rates, foreign exchange rates, money supply, oil prices and gold prices are concurrently major variables that determine the rate of inflation in Indonesia. (Lim & Sek, 2015) This article analyzes, using data collected annually from 1970 to 2011, the variables that influence inflation in two groups of countries. The first is the high-inflation group, and the second is the low-inflation group. Based on Autoregressive Distributed Lag (ARDL) modeling, each factor's short-run and long-run effects on inflation are analyzed. The findings indicate that the GDP and imports had the greatest long-term effect on inflation in economies with low inflation. According to the findings, money supply, national expenditure and Gross domestic product are the long-term determinants of inflation in countries with high inflation. Indeed, none of the factors are significant short-term predictors in high-inflation countries. In nations with low inflation, the supply of money, imports of goods and services and GDP growth are strongly correlated with inflation.

This study (Acquah-Sam, 2017) investigates several primary drivers of Ghana's inflation rate. Multiple linear regression analysis is performed using the conceptual framework and path analysis. The only main factor with a positive and statistically important impact on inflation in Ghana during the study period was the interest rate. In Ghana, factors such as Real GDP, FDI, market capitalization and gross fixed formation have little effect on inflation. The (Loua et al., 2018) study's primary purpose is to assess whether the supply of money, GDP per capita and currency exchange significantly impact the inflation rate in Guinea. According to the Analysis, long-term inflation is affected positively by the supply of money (M2) and the exchange rate. Thus, the effect of GDP per capita on inflation was negative. (Yusof, 2021) The study selected independent variables: government expenditure, exchange rate, unemployment rate and economic growth, whereas the dependent variable is the inflation rate (CPI). This study utilized the OLS method to assess the influence of independent factors on inflation. According to the results, Malaysia's government expenditures, unemployment and exchange rates have a significant relationship with inflation, and the correlation between inflation and economic growth was not statistically significant.

This article (Qurbanalieva, 2013) analyses the key drivers of the price level in Tajikistan using 'autoregressive distributed lags' and Johansen-Juselius cointegration models. This Analysis demonstrates that the currency exchange rate, worldwide wheat prices, labor supply levels and global oil prices affect the long-term price level. However, in the short term, only the world wheat price and labor supply have a significant impact. In the case of demand-pull inflation, the long-term GDP gap, the inflow of remittances and real wages are determined endogenously by the economy, as they have a major impact on the price level. In the short run, the price level is governed by the GDP gap, the influx of remittances, broad money, government expenditure and real wages. The (Durguti, et al., 2021) study will evaluate the effect of remittances, exports, foreign direct investments, GDP growth and imports on the Western Balkans' inflation rate. Besides FDI having a minor impact, other variables affect the inflation rate in the short run. Moreover, the Arellano–Bover/Blundell–Bond estimation demonstrates that GDP growth, FDI and imports positively affect the inflation rate. In opposite, exports and worker remittances have a negative impact on the economy.

Qayyum (2006) Uses correlation analysis to analyze the relationship between Pakistan's money supply (M2) and inflation from 1960 to 2006. His studies revealed a significant relationship between monetary expansion and inflation rate. The study also confirmed that Pakistan's inflation resulted from an expansionary monetary policy. (Biresaw, 2014) uses quarterly data from 1998 and 2010 to

determine the causes and implications of Ethiopia's inflation and uses GDP growth, money supply (M2) and devaluation rate as independent variables for the analysis. Empirical evidence indicated that inflation and money supply were mutually causative; nevertheless, there was a causality relationship between the devaluation rate and the price of oil to the inflation rate. (Gyebi et al., (2013), utilizing the method of ordinary least squares (OLS), identify the macroeconomic variables that affect inflation in Ghana from 1990 to 2009. Output and money supply were the most significant short-run and long-run inflation predictors during the study period. According to their results, the central bank's independence is essential for preserving price stability and sufficient economic growth.

Using the VAR model, quarterly data from 1990 to 2005 (Mwas, 2006) was used to determine the impact of Tanzania's currency exchange rate on inflation. It was determined that, despite the currency's depreciation, the pass-through of the currency exchange rate to inflation declined in the late 1990s. This was partially due to the economic improvements made throughout the research period. The paper stated that the recent rising imports in the country could result in a higher inflation rate over the medium term; however, the government can remain alert in monitoring the potential effect of global prices on Tanzania's inflation.

Durevall et al., (2001) examined the inflation patterns in Kenya using data from time series between 1974 to 1996. They developed a single-equation error correction model that discovered that global prices and exchange rates have a long-term effect on inflation. In contrast, the money supply (M2) and interest rate impact are short-term.

The factors contributing to inflation in Chad is investigated utilizing quarterly data between 1983: Q1 to 2009: Q3 (Kinda, 2011). Paper analysis was based on a model with a single equation and was reinforced with structural vector auto-regression (SVAR) models that can identify shocks and capture inflation persistence. According to the findings, Chad's leading reasons for inflation are precipitation, foreign prices, government expenditures and currency rate volatility. The agricultural sector mediated the relationship between precipitation and domestic pricing.

Mehrara and Sujoudi (2015) used time series data to investigate the link between money supply (M2), inflation and public expenditure in Iran from 1959 to 2010. Using the Bayesian econometric method, the researchers determined that the factors have no significant impact on the country's inflation rate. In addition, the preceding study by (Shaari et al., 2018), which studied panel data from 1986 to 2014, concluded that government expenditure has no significant effect on inflation. This is due to the massive amount of government subsidies offered to products and services, which cut production costs and changed the aggregate supply curve to the right, potentially resulting in a negative effect of government spending on inflation. So, denying the generally accepted notion that government spending positively affects inflation is possible.

According to (Madesha et al., 2013), inflation and the exchange rate are causally related. They emphasized the exchange rate's impact on inflation. Thus, changes in the currency exchange rate will have an instantaneous impact on the entire economy. In addition, the authors explain the long-term relationship between inflation and the exchange rate. When a country's currency depreciates, the price of imported goods rises.

Chakraborty and Joseph (2017) use a collection of macroeconomic data from 1988 to 2015, the performance of several machine learning methods and benchmark autoregressive models for the U.K. was evaluated. They concluded that all machine learning algorithms generate more forecasts than ordinary autoregressive models. (Garcia et al., 2017) compared machine learning techniques' accuracy to standard autoregressive and random walk models in predicting Brazilian inflation from 2003 to 2015.

They found that almost all machine learning techniques exceeded random walk and autoregressive methods. In addition, they anticipated that the accuracy of prediction of machine learning algorithms changes over inflation prediction horizons. Recent research by (Medeiros et al., 2021) demonstrates the significance of machine-learning approaches in U.S. economic inflation forecasting. Owing to the possibility of nonlinear impacts in the relationship between economic determinants and the inflation rate, they indicated that the random-forest method produces the most accurate results. In Costa Rica, (Rodriguez-Vargas, 2020) discovered that combining machine-learning algorithms outperforms univariate forecasting models across all time horizons.

While studies analyzing the accuracy of various econometric models to predict inflation in developed and developing countries, most studies utilizing machine-learning models have found that these techniques are more accurate. As stated, the literature does not provide an example of an inflation forecasting exercise that applies machine-learning techniques to forecast inflation in the Egyptian economy. The difficult economic situation that Egypt is going through necessitates using machine-learning algorithms to predict the inflation rate and determine its causes accurately. Thus, our aim is determining how to confront this.

### 3. Theoretical framework

Many economists and real-world examples of history have proved that inflation remains a significant barrier to long-term economic development and socioeconomic stability, especially in economies in transition. Most central banks' primary mission is to contain inflationary pressure.

Inflation is a characteristic of every economy in the globe. It is a more intricate natural occurrence within a certain economic system than a rise in general prices causes a drop in market value (Aurangzeb & Haq, 2012).

Despite their antique nature, many modern scholars continue applying archaic concepts in theoretical and empirical scenarios. The theoretical arguments about the dimension approach of the inflation rate, namely cost-push and demand-pull, provide a comprehensive and persuasive synthesis of multiple hypotheses for the short- and long-term causes of inflation. Several rival monetarist schools and their followers provide the foundation for the disagreements surrounding these two unique inflation hypotheses.

#### 3.1. Demand-pull approach:

Classical economics, Keynesians and monetarists initially devised different assumptions to analyze the demand-pull inflationary mechanism. Keynesians say it results from income fluctuations and economic shocks, such as oil price rises and other input factors. In contrast, monetarists believe it is a product of strong aggregate demand and inadequate monetary reactions to the economic situation. The most common and oldest type of inflation characterized by pulled-up inflationary pressures resulting from excess aggregate demand for goods and services, the aggregate supply, is indicated as aggregate demand components. (Qurbanalieva, 2013).

The Keynesian hypothesis explains demand-pull inflation as a positive relationship between output and inflation and a negative relationship with unemployment. Thus, growth in the labor force leads to a rise in aggregate demand, which prompts businesses to hire extra workers to supply the increased demand and increase output. However, due to capacity limits, output growth will eventually become so negligible that prices will rise. The year 1940 saw the formulation of the "inflation gap" concept by John Maynard

Keynes and Arthur Smithies, which is now widely accepted (1942). They make plain and specific references to the consequences of the conflict. The source of inflationary pressure is not surplus or “additional” demand nor any imbalance in interest rates; rather, it is the rise in government expenditures. According to their hypothesis, if wages lag behind inflation, inflation will become a transfer mechanism for which some social class will be required to pay the income to counteract the “inflationary gap.”

The Keynesian theory attributes demand-pull inflation to a production capacity shortfall, an excess of aggregate demand. In contrast, classical economics attributes inflation to fluctuations in the aggregate money supply. They contend that the money supply increase exceeds the economy’s capacity to generate the required goods and services. According to monetarists, inflation is a complete monetary phenomenon caused by rising demand. The Monetarists argue that the money supply is a “dominant, but not exclusive” element that affects both short- and long-term pricing but only short-term output. According to this viewpoint, supply and demand in a competitive economy devoid of externalities are mediated via the price mechanism. When the market establishes prices, the optimal allocation of resources is ensured.

### 3.2. *Cost-push approach:*

Businesses cannot maintain profitability by producing the same quantity of goods if their output costs and productivity rise. Thus, the higher costs are passed on to the final consumers, increasing the overall price. The long-term cost-push effect revives economic stagflation. Input costs have increased due to the lack of raw materials and a significant increase in worldwide prices, especially for oil and petrol. It also arises due to exposure to external shocks, such as the fluctuation of oil prices on the international market and the exchange rate depreciation. The rise in production costs imposes inflationary pressure on businesses since they must boost compensation to obtain a highly qualified labor force. Yet, enterprises can successfully pass on this portion of output costs to consumers by raising output prices (Qurbanalieva, 2013).

A supply-side shock commences a process of persistent inflation in a full employment economy. As the economy nears full employment, the unemployment rate declines, prompting employees and their supporters to seek a wage increase. Companies raise production costs while retaining the markup to prevent this salary increase from undermining earnings. A demand for actual wage testing in wage-price cycles propagates through the indexation process. Eventually, the real salary falls due to increased food costs. Especially in a tiny free market considered a price-taker, the decline in the local currency could affect the price of imported foodstuffs, raw materials and capital goods.

## 4. **Empirical framework**

The level of prices is influenced by monetary and non-monetary elements, as well as local and global ones. This section describes the empirical framework for estimating the Egyptian inflation drivers. In this study, the nine variables gross domestic product growth, government spending (% of GDP), household spending (% of GDP), external trade balance (% of GDP), Fixed capital formation (% GDP), FDI inflow (% of GDP), exchange rates, GDP per capita and money supply (M2) are identified as the major determinants of inflation rate in Egypt. The following equation shows the variables used in analyzing the determinants of inflation in Egypt:

$$INF = F(GDPg, GS, HS, ETB, FCF, FDI, ER, GDPC, M2)$$

Where;

INF= inflation rate based on the consumer price index

GDPG= gross domestic product growth

GS = government spending (% of GDP)

HS = household spending (% of GDP)

ETB = external trade balance (% of GDP)

FCF = Fixed capital formation (% of GDP)

FDI = foreign direct investment inflow (% of GDP)

ER = exchange rates

GDPC = GDP per capita

M2 = money supply

## 5. Data and methodology

### 5.1. data

**Table 1.** Description and descriptive statistics of the variables.

| Variables                               | abbreviation | Source of data                  | mean  | mode       | median | dispersion | min    | max   |
|---|--------------|---------------------------------|-------|------------|--------|------------|--------|-------|
| Inflation rate (Δ consumer price index) | INF          | World bank                      | 11.8  | 2.26       | 11.07  | 0.52       | 2.26   | 29.5  |
| GDP growth (%)                          | GDPG         | World bank                      | 5.28  | 1.12       | 4.92   | 0.44       | 1.12   | 13.27 |
| Government spending (% GDP)             | G.S.         | World bank                      | 12.51 | 7.66       | 11.42  | 0.25       | 7.66   | 23.43 |
| Household spending (% GDP)              | H.S.         | World bank                      | 74.26 | 63.09      | 73.76  | .092       | 63.09  | 89    |
| External trade balance (% GDP)          | ETB          | World bank                      | -8.21 | -17.8<br>5 | -8.25  | -0.46      | -17.85 | -1.35 |
| Fixed capital formation (% GDP)         | FCF          | World bank                      | 21.37 | 12.02      | 19.55  | 0.28       | 12.02  | 33.11 |
| FDI inflow (% GDP)                      | FDI          | World bank                      | 2.37  | -0.2       | 1.7    | 0.84       | -0.2   | 9.34  |
| Exchange rates                          | E.R.         | The Central Bank of Egypt (CBE) | 5.33  | 0.7        | 3.39   | 1.03       | 0.39   | 24.7  |
| GDP per capita                          | GDPC         | World bank                      | 2.9   | -1.28      | 2.37   | 0.77       | -1.28  | 10.78 |
| Money supply growth (%)                 | M2           | World bank                      | 18.93 | 5.66       | 18.5   | 0.46       | 5.66   | 51.41 |

Source: Compiled by the author.

We use big data containing economic variables of the Egyptian economy. To accurately analyze the determination of inflation, the paper depends on a large time series from 1976 to 2022. The extensive data contains sets of variables including GDP growth, government spending (% GDP),

household spending (% GDP), external trade balance (% GDP), Fixed capital formation (% GDP), FDI inflow (% GDP), exchange rates, GDP per capita and money supply. Table 1 contains the list, definition, description, descriptive statistics and data sources for the variables used.

## 5.2. Methodology

We use machine learning algorithms like KNN, R.F., SVM, G.B., D.T. and ANN to analyze the determination of Egyptian inflation in the next paragraphs, showing how it works. Next, the section determines the algorithms for the analysis and will depend on them for forecasting and getting features important to make accurate economic policy to face the inflation problem.

### 5.2.1. Random forest

The Random Forest (R.F.) machine learning approach is an ensemble technique with regression and classification capabilities. Instead of relying on a single decision tree to determine the final output of the model, the approach combines multiple trees to reduce the model's variance. During the training of each decision tree, additional factors were added, affecting the decisions at the node level. Due to the large share market size, the data's noise can sometimes be substantial, which can cause trees to evolve in radically unexpected ways. The following day's closing share price for a certain firm is calculated using training factors to minimize forecasting errors (Vijh et al., 2020).

### 5.2.2. SVM

To classify two data sets, SVMs were initially presented as an algorithm that maximizes the distance or margin between each class's closest point and its separating hyperplane. This definition of SVM is accurate if a linear border can differentiate between classes. If the classes cannot be distinguished, the hyperplane limitations can be changed to permit observations on either side of the hyperplane or even within the boundary. Nonlinear boundaries suggest that an SVM based on a linear hyper-plane cannot classify the observations accurately. The data are projected into a higher-dimensional space with a linear boundary using kernel-based basis functions to address this issue. The kernel function tuning parameters enable SVM models to model the proper nonlinear boundary conditions (Chen, 2020).

### 5.2.3. KNN

Neighboring datasets are called "neighbors" because they are geographically close. Data, distance and vector modules are utilized to compare the similarities of datasets. With the provided dataset, a constant "k" is employed to locate the closest point in the distance. Next, the distance between each pair of data points is plotted and calculated, typically using Euclidean geometry.

Focusing on data proximity (Mohamed et al., 2021) this dataset contains a wide array of values and their vicinity. A grid of ascending distance values is presented. The closest distance index "k" was chosen to sort the array, and the location of that index was utilized to sort the array. A wide variety of datasets exist, all dispersed across the globe. It is more feasible due to the distribution of identical items.



Each division contains little data pieces, and the results are determined by their proximity. It is a straightforward algorithm and its operation is evident. These datasets are called “nonlinear” because no initial assumptions were made about the data.

It is a useful and flexible method that may be used to classify and regress datasets. The yield factor is the most precise and effective method for producing results.

#### 5.2.4. Gradient boosting

The original boosting technique, Adaboost (Freund and Schapire, 1997), identifies a hypothesis with a low prediction error relative to a specified distribution across the training data. Gradient boosting, the generalization of Adaboost, employs this approach. For instance, Freund and Schapire exemplified their algorithm using a horse-betting scenario in which a bettor seeks the most probable winner. Before placing a wager, gamblers should consult professionals to increase their odds of success. This collection of data from multiple professionals resembles an incompetent classification team. A training set is prepared for each Adaboost expert’s viewpoint. After iteration, the weights of the training sets are updated to increase the weights of incorrectly categorized samples and decrease the weights of correctly classified samples.

The current weak learner of Adaboost selects a weak hypothesis from the entire set of weak hypotheses, as opposed to only the weak hypotheses identified in each boosting iteration. While looking for all hypotheses in a broad area might be laborious, selecting students who cover most of the set is generally preferable (He et al., 2019).

#### 5.2.5. Neural network

The brain’s biological neural networks are the basis for artificial neural networks (ANNs). ANNs are superior to mathematical algorithms for pattern recognition and matching, grouping and classification. There are benefits and drawbacks to each sort of algorithm.

The first ANN was constructed utilizing a fundamental understanding of brain connections. Frank Rosenblatt, a neurobiologist, hypothesized that supervised learning might modify the artificial connections between neurons in the Mark I Perceptron machine, lowering the disparity between real and expected outputs. Due to the difference in traveling back across the network, the weights of the links can be modified. As a result of this disparity, the network has the information required to enhance learning outcomes. (Dell’Aversana, 2019) uses a training dataset to predict the outcome.

#### 5.2.6. Decision tree

This is a tree-based, if-then machine learning technique with the moniker “if-then” since it behaves like an if-statement. The root node of the decision tree is produced initially, followed by its children. The information is categorized depending on the attributes of nodes that represent decision points. The branches linking the nodes at different levels reflect alternative options, which are determined by examining the attributes of each node (Sen & Engelbrecht, 2021).

## 6. Empirical results

### 6.1. Models evaluation

Mean absolute error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Coefficient of determination (R<sup>2</sup>) are the most common metrics employed in regression analysis to evaluate forecast error rates and highest accuracy.

The MAE reflects the difference between actual and expected values, as determined by averaging the absolute difference throughout the full data set; it is computed using the following Equation (1):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

The MSE represents the difference between actual and predicted values and is derived by quadrupling the average difference across the full data set. It is computed using the following Equation (2):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2 \quad (2)$$

The RMSE is the error rate multiplied by the square root of the MSE. It is calculated using the following Equation (3):

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2} \quad (3)$$

The R<sup>2</sup> measures how well the values fit relative to the initial values. The greater the value of R<sup>2</sup>, the better the model. It is calculated using the Equation (4): The proportion of values between zero and one.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

### 6.2. Determine the accuracy algorithms

Table 2 present the accuracy results for the employed algorithms after processing them using Python.

**Table 2.** Machine-learning algorithm's accuracy in predicting inflation, 1976–2022.

| Model             | MSE   | RMSE  | MAE   | R <sup>2</sup> |
|-------------------|-------|-------|-------|----------------|
| Gradient Boosting | 0.039 | 0.196 | 0.161 | 0.99           |
| Decesion Tree     | 3.58  | 1.89  | 1.23  | 0.90           |
| Random Forest     | 4.62  | 2.15  | 1.59  | 0.87           |
| Neural Network    | 13.83 | 3.71  | 2.94  | 0.63           |
| KNN               | 21.65 | 4.65  | 3.70  | 0.42           |
| SVM               | 24.08 | 4.9   | 3.72  | 0.36           |

Source: Compiled by the author.

According to Table 2, the Gradient boosting algorithms provide the best accurate predictions, followed by decision tree and random forest, also with high accuracy and reasonably.

### 6.3. prediction performance:

Table 3 and Figure 1 show that the actual and expected values are almost identical, which indicates the accuracy of the G.B. algorithm's performance in forecasting, which results in successful economic policies.

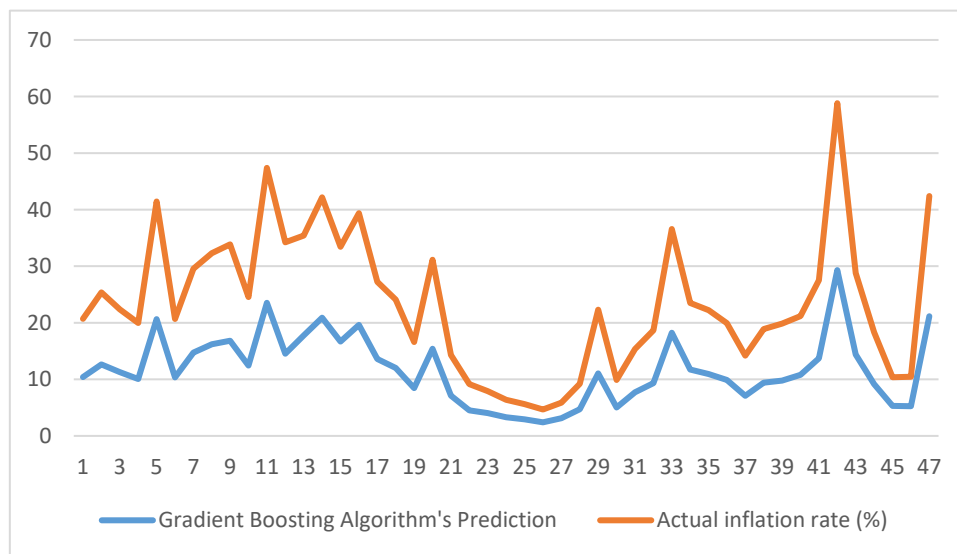
**Table 3.** Gradient boosting algorithm prediction for Egypt's inflation rate.

| year | Actual inflation rate (%) | Gradient Boosting Algorithm's Prediction |
|------|---------------------------|--|
| 1976 | 10.32                     | 10.39                                    |
| 1977 | 12.73                     | 12.64                                    |
| 1978 | 11.08                     | 11.28                                    |
| 1979 | 9.90                      | 10.07                                    |
| 1980 | 20.82                     | 20.63                                    |
| 1981 | 10.32                     | 10.35                                    |
| 1982 | 14.82                     | 14.74                                    |
| 1983 | 16.08                     | 16.21                                    |
| 1984 | 17.04                     | 16.82                                    |
| 1985 | 12.11                     | 12.43                                    |
| 1986 | 23.86                     | 23.52                                    |
| 1987 | 19.69                     | 14.53                                    |
| 1988 | 17.66                     | 17.76                                    |
| 1989 | 21.26                     | 20.91                                    |
| 1990 | 16.76                     | 16.68                                    |
| 1991 | 19.75                     | 19.63                                    |
| 1992 | 13.64                     | 13.59                                    |
| 1993 | 12.09                     | 12.01                                    |
| 1994 | 8.15                      | 8.45                                     |
| 1995 | 15.74                     | 15.4                                     |
| 1996 | 7.19                      | 7.12                                     |
| 1997 | 4.63                      | 4.52                                     |
| 1998 | 3.87                      | 4.04                                     |
| 1999 | 3.08                      | 3.30                                     |
| 2000 | 2.68                      | 2.93                                     |
| 2001 | 2.27                      | 2.40                                     |
| 2002 | 2.74                      | 3.12                                     |
| 2003 | 4.51                      | 4.7                                      |
| 2004 | 11.27                     | 11.04                                    |
| 2005 | 4.87                      | 5.03                                     |
| 2006 | 7.64                      | 7.72                                     |
| 2007 | 9.32                      | 9.33                                     |

*Continued on next page*

| year | Actual inflation rate (%) | Gradient Boosting Algorithm's Prediction |
|------|---------------------------|--|
| 2008 | 18.32                     | 18.25                                    |
| 2009 | 11.76                     | 11.73                                    |
| 2010 | 11.27                     | 10.95                                    |
| 2011 | 10.06                     | 9.88                                     |
| 2012 | 7.11                      | 7.09                                     |
| 2013 | 9.47                      | 9.42                                     |
| 2014 | 10.07                     | 9.77                                     |
| 2015 | 10.37                     | 10.8                                     |
| 2016 | 13.81                     | 13.71                                    |
| 2017 | 29.51                     | 29.31                                    |
| 2018 | 14.40                     | 14.4                                     |
| 2019 | 9.15                      | 9.16                                     |
| 2020 | 5.04                      | 5.32                                     |
| 2021 | 5.21                      | 5.24                                     |
| 2022 | 21.26                     | 21.14                                    |

Source: Compiled by the author.



Source: Compiled by the author.

**Figure 1.** Prediction performance.

#### 6.4. The gradient boosting feature importance

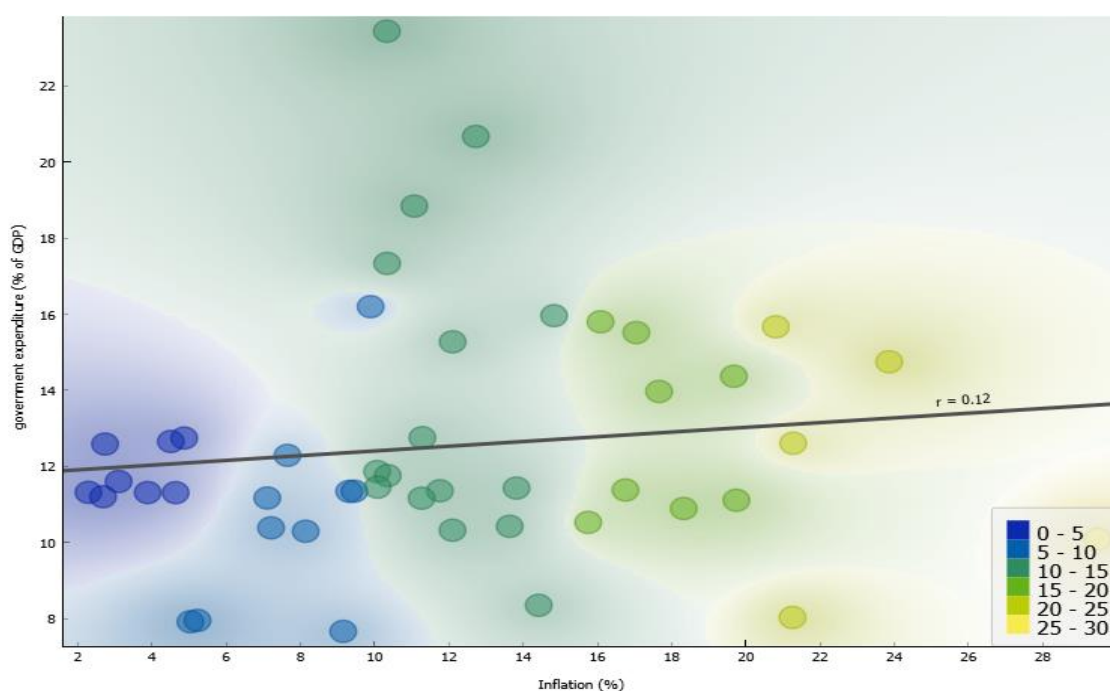
**Feature Importance:** Although these techniques are typically employed for prediction, learning which variables have the greatest impact on a model may be accomplished by analyzing their feature importance. The outcomes of this analysis are displayed in Table 4.

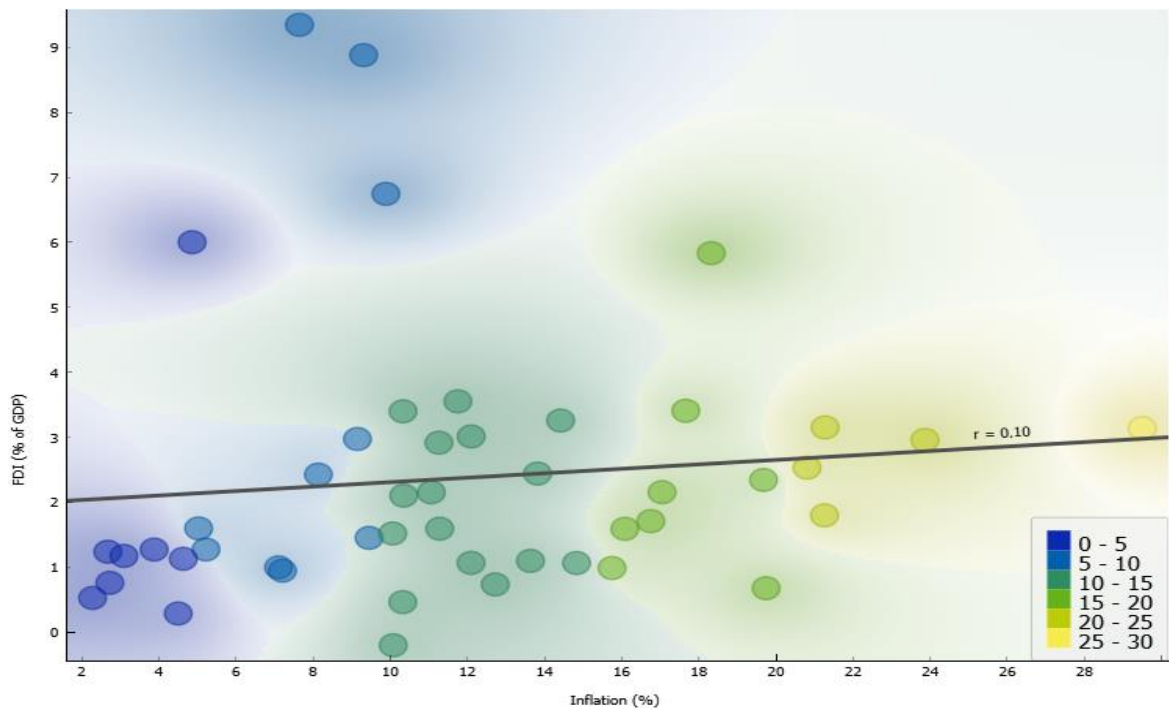
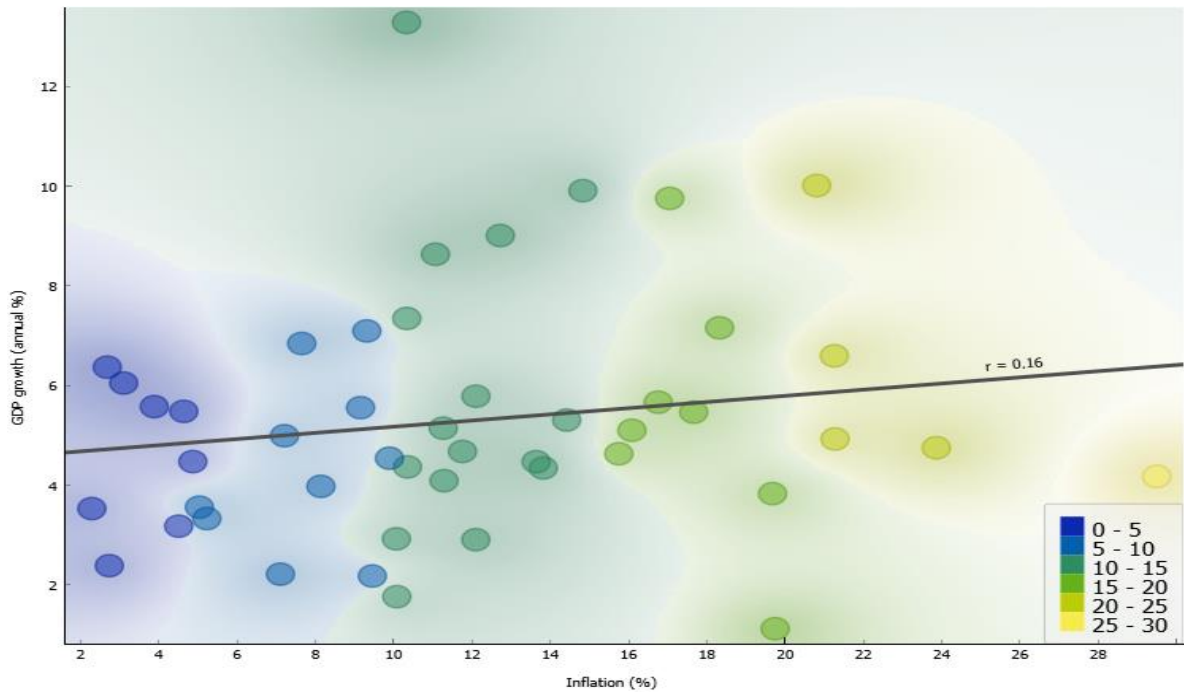
**Table 4.** Feature importance indicators.

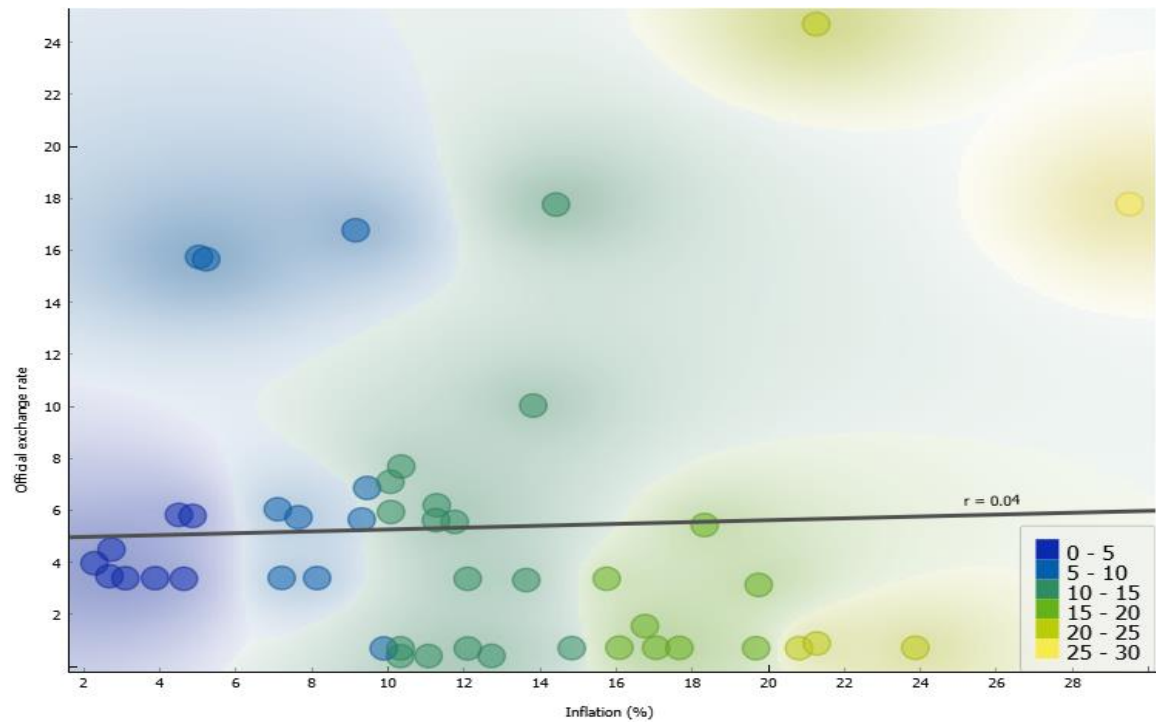
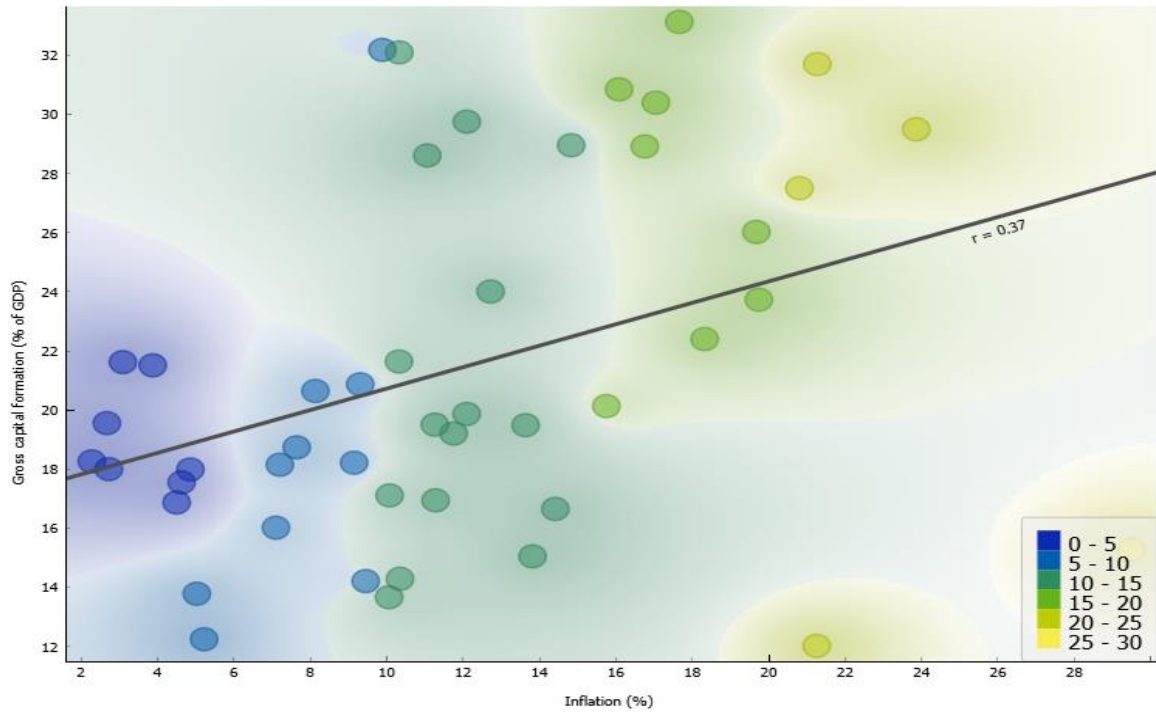
| variables | Percentage (%) |
|-----------|----------------|
| ER        | 30.5           |
| FCF       | 24.5           |
| GS        | 12.3           |
| FDI       | 9.1            |
| GDPC      | 5.3            |
| M2        | 5              |
| GDPG      | 4.9            |
| HS        | 4.8            |
| ETB       | 3.7            |

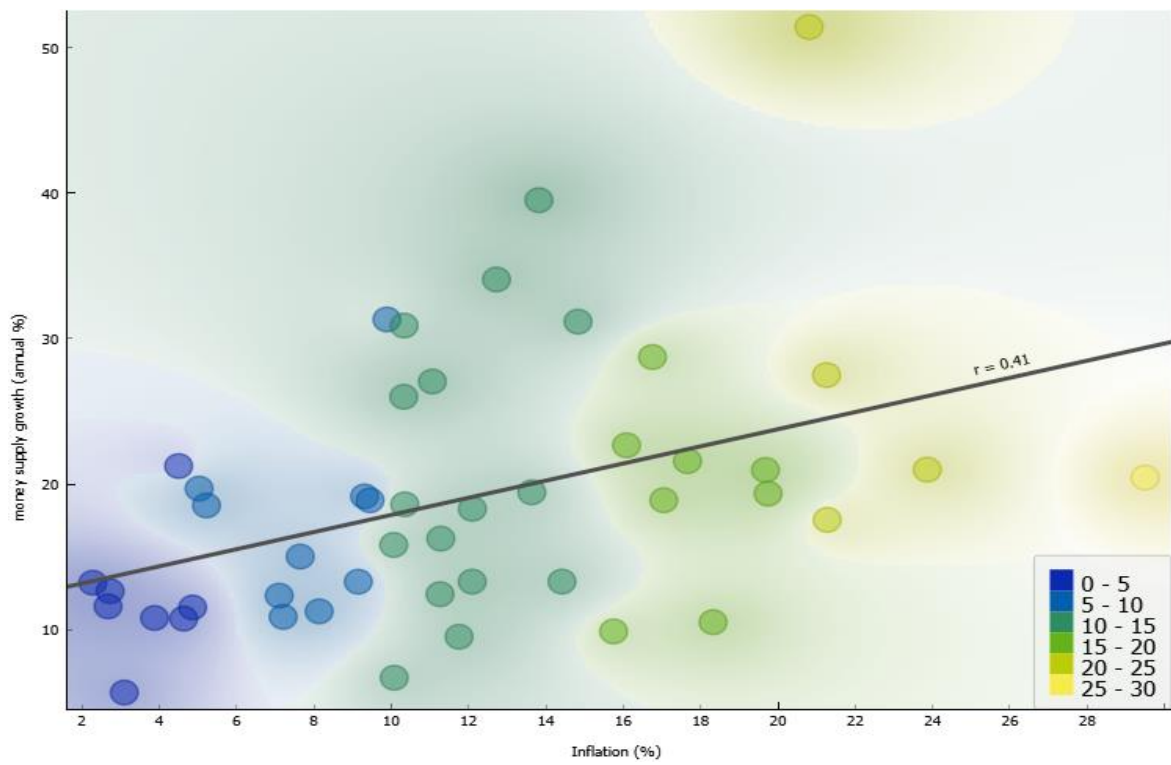
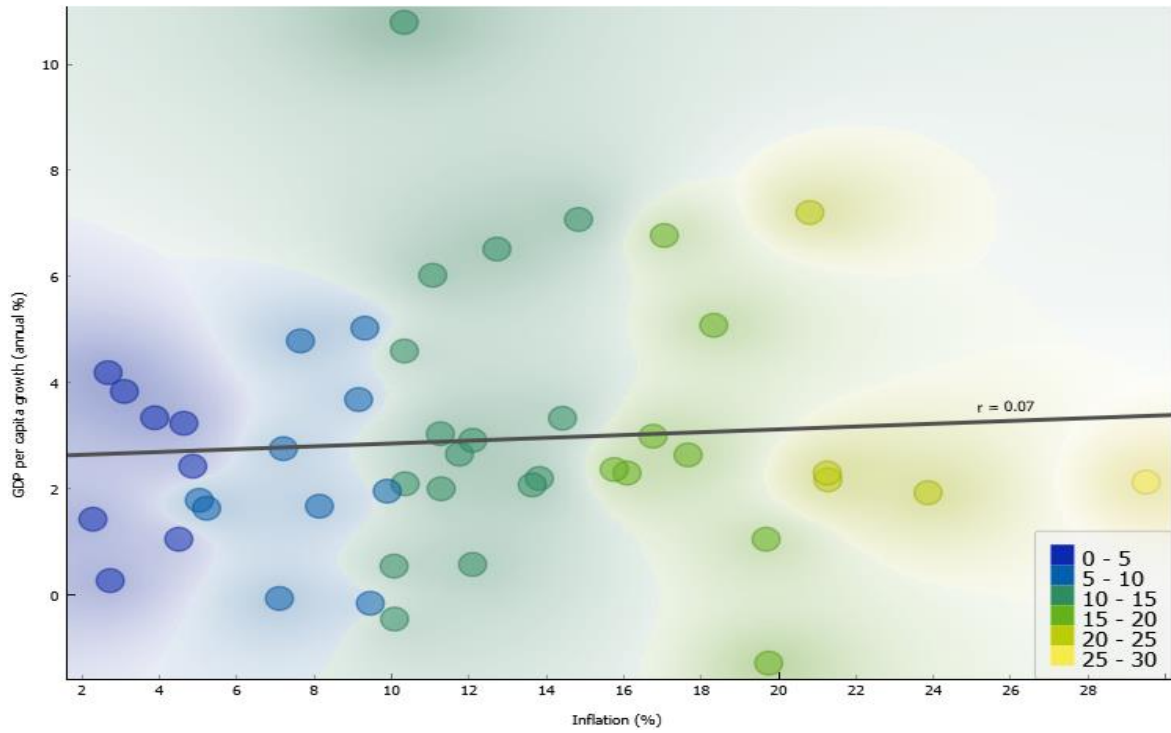
From Table 4, the most affect on Egypt's inflation rate is the exchange rate at 30.5 %, followed by the fixed capital formation by 24.5%, government spending by 12.3%, FDI inflow by 9.1%, GDP per capita by 5.3%, money supply by 5%, GDP growth by 4.9%, household expenditure by 4.8% and external trade balance by 3.7%. Hence, to build a successful economic policy to address inflation, we must focus on the four major influences of the exchange rate, fixed capital formation and government spending.

Now, we are determining the relationship between the dependent and independent variables. Is it positive or negative? To accurately determine this, the following scatter plot in Figure (2) can show these relations.

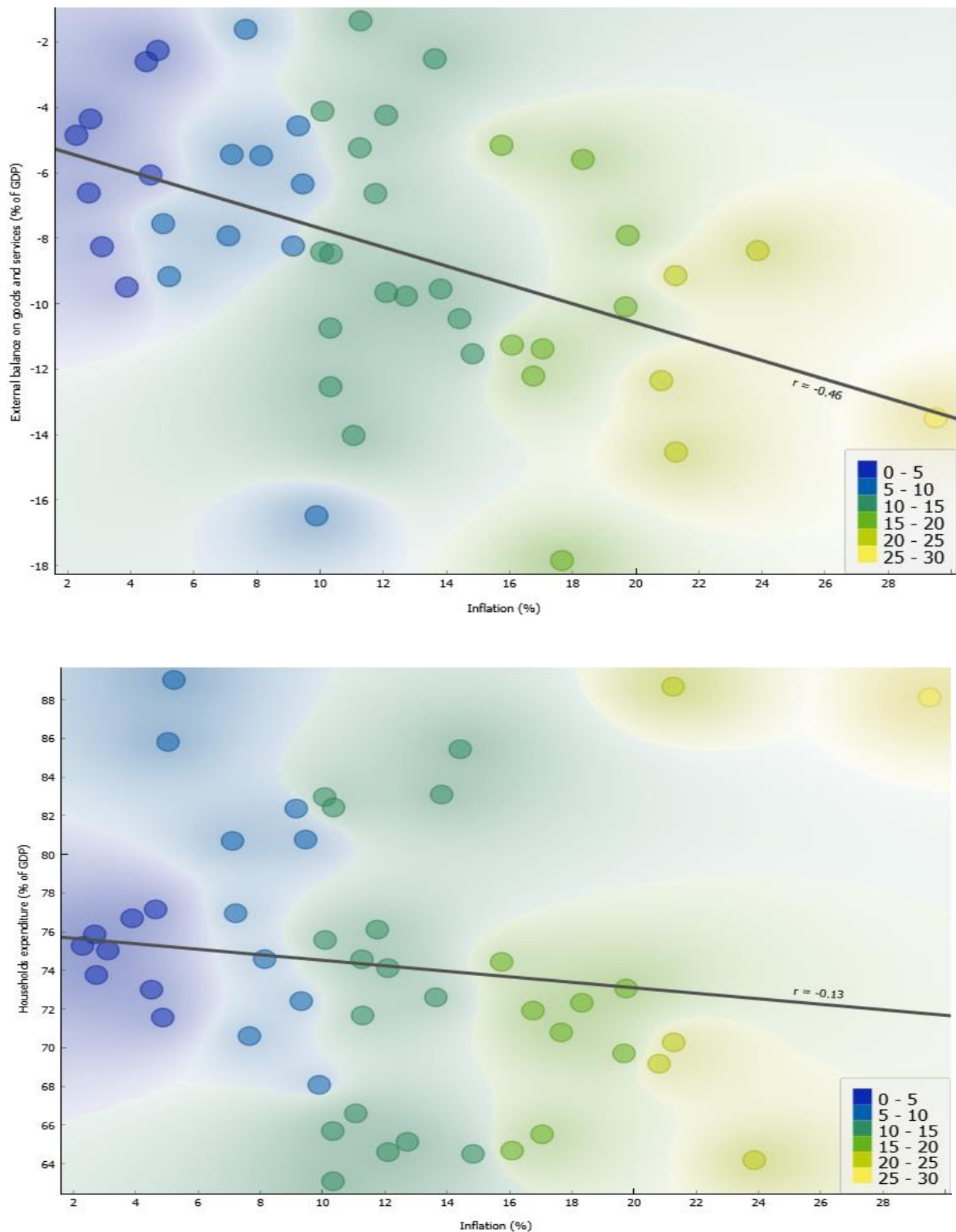












Source: Compiled by the author.

**Figure 2.** Scatter plot for the relationship between inflation rate and other variables.

The scatter plot shows a positive relationship between the inflation rate and GDP growth, gross capital formation (investment expenditure), government expenditure, FDI, money supply, exchange

rate and GDP per capita. Furthermore, the relationship between inflation, household expenditure and the external trade balance is negative.

## 7. Conclusions

Inflation is considered one of the most difficult problems that the Egyptian economy is facing, so predicting it and identifying its most important determinants is essential for building a successful economic policy. Therefore, six machine learning algorithms (Support vector machine, Tree, K-nearest neighbor, Random Forest, Neural Network, Gradient boosting and decision tree) were used to determine the most accurate and efficient one to achieve the desired purpose. It turns out that the gradient enhancement algorithm is the one that will fulfill the purpose with an  $R^2$  (0.99), MSE (0.039), RMSE (0.19) and MAE (0.16).

Many variables have been relied upon, some of which are associated with the inflation rate in a positive relationship, such as GDP growth, government expenditure, gross capital formation (investment expenditure), FDI, exchange rate, GDP per capita and money supply. In addition, others have a negative relationship with it, like household expenditure and the external trade balance. The most important determinant of the inflation rate in Egypt is the exchange rate at 30.5 %, then the fixed capital formation by 24.5%, government spending by 12.3%, FDI inflow by 9.1%, GDP per capita by 5.3%, money supply by 5%, GDP growth by 4.9%, households' expenditure by 4.8% and external trade balance by 3.7%.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Conflict of interest

The authors declare no conflict of interest.

## References

- Abd El-Aal MF, Algarni A, Fayomi A, et al. (2021) Forecasting foreign direct investment inflow to Egypt and determinates: Using machine learning algorithms and ARIMA model. *J Adv Transport* 2021: 1–7. <https://doi.org/10.1155/2021/9614101>
- Acquah-Sam E (2017) Influencers of Inflation in Ghana. *Eur Sci J* 13: 140–158.
- Aurangzeb AUH, Haq UA (2012) Determinants of inflation in Pakistan. *J Manag Soc Sci* 2: 89–96.
- Baybuza I (2018) Inflation forecasting using machine learning methods. *Russ J Money Financ* 77: 42–59. <https://doi.org/10.31477/rjmf.201804.42>
- Biresaw TT (2014) Determinant and impacts of dynamic inflation in Ethiopia. Master's thesis, Norwegian University of Life Sciences, As.
- Chakraborty C, Joseph A (2017). Machine learning at central banks.
- Chen, LP (2020) Using machine learning algorithms on prediction of stock price. *J Model Optim* 12: 84–99. <https://doi.org/10.32732/jmo.2020.12.2.84>

- DELL'AVERSANA P (2019) Comparison of different Machine Learning algorithms for lithofacies classification from well logs. *B Geofis Teor Appl* 60. <https://doi.org/10.4430/bgta0256>
- Durevall D, Ndung'u NS (2001) A dynamic model of inflation of Kenya, 1974–1996. *J Afr Econ* 10: 92–125.
- Durguti E, Tmava Q, Demiri-Kunoviku F, et al. (2021) Panel estimating effects of macroeconomic determinants on Inflation: Evidence of Western Balkan. *Cogent Eco Financ* 9: 1942601. <https://doi.org/10.1080/23322039.2021.1942601>
- Freund Y, Schapire RE (1997) A decision-theoretic generalization of online learning and an application to boosting. *J Comput Syst Sci* 55: 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman M (1968) Dollars and deficits: inflation, monetary policy, and the balance of payments (No. 332.4/F91d).
- Garcia MG, Medeiros MC, Vasconcelos GF (2017) Real-time inflation forecasting with high dimensional models: the case of Brazil, *Int J Forecasting* 33: 679–693. <https://doi.org/10.1016/j.ijforecast.2017.02.002>
- Garcia-Bellido J, Morales ER (2017) Primordial black holes from single field models of inflation. *Phys Dark Universe* 18: 47–54. <https://doi.org/10.1016/j.dark.2017.09.007>
- Groen JJ, Paap R, Ravazzolo F (2013) Real-time inflation forecasting in a changing world. *J Bus Econ Stat* 31: 29–44. <https://doi.org/10.1080/07350015.2012.727718>
- Gyebi F, Boafo GK (2013) Macroeconomic determinants of inflation in Ghana from 1990–2009. *Int J Bus Soc Res* 3: 81–93.
- He Z, Lin D, Lau T, et al. (2019) Gradient boosting machine: a survey. arXiv preprint [arXiv:1908.06951](https://arxiv.org/abs/1908.06951).
- Iversen J, Laseen S, Lundvall H, et al. (2016) Real-time Forecasting for Monetary Policy Analysis: The Case of Sveriges Riksbank. *Riksbank Res Pap Ser* 142. <https://doi.org/10.2139/ssrn.2780417>
- Kinda MT (2011) *Modeling inflation in Chad*. International Monetary Fund. <https://doi.org/10.5089/9781455221011.001>
- Lim YC, Sek SK (2015) An examination of the determinants of inflation. *J Econ Bus Manage* 3: 678–682. <https://doi.org/10.7763/JOEBM.2015.V3.265>
- Loua A, Wan Rong Y, Koivogui SK, et al. (2018) Macroeconomic Determinants of Inflation: Evidence from the Republic of Guinea. *Res J Financ Account* 9: 13–21.
- Madesha W, Chidoko C, Zivanomoyo J (2013) Empirical test of the relationship between exchange rate and inflation in Zimbabwe. *J Econ Sustain Dev* 4: 52–58.
- Medeiros MC, Vasconcelos GF, Veiga Á, et al. (2021) Forecasting inflation in a data-rich environment: the benefits of machine learning methods. *J Bus Econ Stat* 39: 98–119. <https://doi.org/10.1080/07350015.2019.1637745>
- Mehrara M, Sujoudi A (2015) The relationship between money, government spending, and inflation in the Iranian economy. *Int Lett Soc Humanistic Sci* 51: 89–94. <https://doi.org/10.18052/www.scipress.com/ILSHS.51.89>
- Monteforte L, Moretti G (2013) Real-time forecasts of inflation: The role of financial variables. *J Forecast* 32: 51–61. <https://doi.org/10.1002/for.1250>
- Mwase N (2006) An empirical investigation of the exchange rate pass-through to inflation in Tanzania. 2006: 34. <https://doi.org/10.2139/ssrn.920242>
- Qayyum A (2006) Money, inflation, and growth in Pakistan. *Pakistan Dev Rev* 45: 203–212. <https://doi.org/10.30541/v45i2pp.203-212>

- Qurbanalieva N (2013) An empirical study of factors affecting inflation in the Republic of Tajikistan. *Theo Pract Res Econ F* 4: 229–246
- Rodríguez-Vargas A (2020) Forecasting Costa Rican inflation with machine learning methods. *Lat Am J Cent Bank* 1: 100012. <https://doi.org/10.1016/j.latchb.2020.100012>
- Sen J, Mehtab S, Engelbrecht A (2021) *Machine Learning: Algorithms, Models, and Applications*. BoD—Books on Demand. <https://doi.org/10.5772/intechopen.94615>
- Shaari MS, Ahmad TST, Razali R (2018) Tourism led-inflation: A case of Malaysia. In MATEC Web of Conferences. 150: 06026. EDP Sciences. <https://doi.org/10.1051/mateconf/201815006026>
- Vijh M, Chandola D, Tikkiwal VA, et al. (2020) Stock closing price prediction using machine learning techniques. *Procedia Comput Sci* 167: 599–606. <https://doi.org/10.1016/j.procs.2020.03.326>
- Yadav O, Gomes C, Kanojiya A, et al. (2019) Inflation prediction model using machine learning. *Int J Inform Comput Sci* 6: 121–128.
- Yolanda Y (2017) Analysis of factors affecting inflation and its impact on the human development index and poverty in Indonesia. <https://doi.org/10.35808/ersj/873>
- Yusof N, Nin LF, Kamal HKM, et al. (2021) Factors that influence the inflation rate in Malaysia. *Int J Acad Res Bus Soc Sci* 11: 626–637. <https://doi.org/10.6007/IJARBSS/v11-i9/10838>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)