Research Article

Forecasting Foreign Direct Investment Inflow to Egypt and Determinates: Using Machine Learning Algorithms and ARIMA Model

Mohamed F. Abd El-Aal,1 Ali Algarni,2 Aisha Fayomi,3 RAahayu Abdul Rahman,4 and Khudir Alrashidi5

1Higher Institute of Commercial Sciences, Department of Economic, Mahla, Egypt
2Faculty of Science Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia
3Faculty of Science Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia
4Faculty of Accountancy, Universiti Teknologi MARA, Seri Iskandar 32610, Perak, Malaysia
5Faculty of Science Department of Statistics, King Abdulaziz University, Jeddah, Saudi Arabia

Correspondence should be addressed to Mohamed F. Abd El-Aal: fawzi_1988@yahoo.com and Ali Algarni: ahalgarni@kau.edu.sa

Received 12 October 2021; Revised 8 November 2021; Accepted 18 November 2021; Published 6 December 2021

Academic Editor: Muhammad Arif

Copyright © 2021 Mohamed F. Abd El-Aal et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study aims to determine the primary determinant of FDI inflow to Egypt using machine learning algorithms and the ARIMA model and get an accurate prediction of FDI inflow to Egypt during the current decade (2020–2030) and approved that the gradient boosting model is the most accurate algorithms. Also, we find stability in economic indicators in Egypt during the current decade using the ARIMA model. The last step approved that the primary determinant of FDI inflow to Egypt is the Human Development Index, followed by population size, gross domestic product per capita, lending rate, and gross domestic product value.

1. Introduction

Developed and developing countries seek to attract foreign direct investment (FDI) because of their positive effects on their growth levels by increasing production, rising employment, increasing exports, and transferring modern technology. Therefore, countries compete to provide the advantages and facilities necessary to attract this type of investment. Today, the flow of foreign direct investment from the North to the South (from developed countries to developing countries) is not limited to, but flows from the South to the South (from developing countries to developing countries), for example, the People’s Republic of China. It plays a significant role in developing countries, particularly in eliminating their domestic resource gap, as most countries suffer from a gap in their domestic savings.

In the last two decades, FDI flows to Egypt have been essentially random. During them, the value of foreign direct investment flows reached the top in the years 2006 and 2007, with a value of 10 and 11.5 billion dollars, respectively. Then, its value gradually decreased in 2008, 2009, and 2010 to reach 9.5, 6.7, and 6.3 bn$, respectively. However, the value of foreign direct investment reached a negative value in 2011 due to the January 25 revolution. However, the flows were regulated from 2013 to 2019, but they did not reach the values of popularity in the years 2006 and 2007 “World Bank Data [1].”

According to the geographical distribution of FDI inflow in Egypt 2017/2018, indicating the European Union with the most significant contribution to Egypt FDI inflow by 60.4% from the total amount of FDI inflow, then the United States of America by 17%, then Arab countries by 14.8%, and other countries by 7.9%. In the same fiscal year, the sectoral
distribution indicated that the oil sector gets most of FDI inflow by 67.3%, the services’ sector by 11.2%, the manufacturing sector by 10%, construction sector by 4.5%, activities unallocated 6.9%, and agriculture sector by 0.1% “central bank of Egypt, 2017 [2].”

To select the most accurate and efficient model, given the importance of predicting foreign direct investment and its determinants, the researcher used machine learning algorithms to accurately determine its value to enable the decision maker to set sound economic policy. In a first step, we predict foreign direct investment using machine learning algorithms, especially random forest, support vector machine (SVM), logistic regression, na"ıve Bayes, k-nearest neighbors (K.N.N.), neural network, and gradient boosting algorithms. Then, in a second step, the autoregressive integrated moving average model (ARIMA) is used to analyze the time series of the independent variables. Then, in a third step, we use the predictions of the independent variables deduced from the ARIMA model, use them in the most accurate algorithm to obtain the future values of foreign direct investment inflow into Egypt and determine the most important independent variables precisely for it.

2. Literature Review

Singh [3] used machine learning algorithms, especially random forest and gradient boosting. This study used a large number of variables such as the total amount of FDI, the amount of FDI inflow, production, agglomeration of industries, and G.D.P per capita, the number of the students in higher school education, urbanization, health services, employment, number of active firms, wages, and grouping of the manufacturing industry. The result of this study is the production variable is the most influential determinate to FDI inflows.

3. The Determinates of FDI Inflow:

Empirical Evidence

The work of Dunning [4–6] (1993, 2000, and 2008) is a significant reference when studying foreign direct investment motives; he determines it to four motives: first, market seeking FDI search about local and regional markets, wherein these markets get a production facility and tariff jumping, second, resource seeking the motives of FDI to obtain natural resources and raw materials, often not available in its home countries, third, efficiency-seeking FDI search about the country that more openness and part of the world and FDI get benefits of economy of scale results of specialization, and fourth, strategic assets’ seeking to achieve long-term strategic goals, and companies often acquire the assets of foreign firms. The goal of this motive is to get the benefits of shared ownership and reduce the competition.

“UNCTAD (1998)”: according to the world investment report issued by the United Nations conference on trade and development, the FDI inflow factors are determined, and these factors are classified into three groups: political, business, and economic factors. Tiskat et al. [7] used gross domestic product (G.D.P.), G.D.P. per capita, G.D.P. growth, public investment (% G.D.P.), corporate ratio tax (% G.D.P.), exchange rate, and lending rate as a determinate of FDI. Oecd [8] used G.D.P., exports’ value, imports’ value, exchange rate, and globalization index to determine FDI. In 2018, Asimamah et al [9] used inflation rate, interest rate, exchange rate, real G.D.P., electricity production, and telephone usage as determinates of FDI. Hintosova [10] used market size, economic stability, and innovation as determinants of FDI in Visegrad countries. Songu et al. [11] used G.D.P., G.D.P. growth, G.D.P. per capita, population growth, population total, national resources of G.D.P., and human development index.

Based on previous studies in this paper, we used the market size (real G.D.P., G.D.P. per capita, and population total), trade openness (trad % G.D.P.), macroeconomic stability (unemployment rate and inflation rate), finance cost (lending interest rate), and human capital (human development index) as determinates’ FDI inflow in Egypt. Indicators that are difficult to predict, such as political stability and fixed values such as labor cost (the minimum wage is fixed in Egypt), have been excluded and do not affect the models used.

4. Methodology

This paper used random forest, SVM, logistic regression, Naïve Bayes, K.N.N., neural networks, and gradient boosting models. All models have monitored machine learning models, which means that they analyze data using training data and then create a data prediction function.

Seven models are used to determine their accuracy and the accuracy model for FDI prediction. We use the ARIMA model for time-series analysis.

The World Bank dataset was used to collect real G.D.P., G.D.P. per capita, population, trade volume (% G.D.P.), inflation rate, unemployment rate, lending interest rate, and human development index data from 1990 to 2019. The machine learning algorithms utilized in this study were written in Python using the Scikit-Learn package.

5. Models

5.1. Random Forest. In a random forest, many different decision trees make up the forest. The binary outcome variable is predicted using a classification decision tree instead of a serial number. These two forms of decision trees split the data into two categories at each decision point similarly. There is a yes or no decision at every node. Is $x$ larger than 5? Based on the response, the data are then partitioned into smaller groups. In the next step, new explanatory variables are added to repartition the data once again. As a result, the first explanatory variable chosen may account for the most significant data separations. That smaller bucket’s model forecast is the mean value of the separated data bucket. An overfitted decision tree can develop when a decision tree has too many partitions. This results in a model that performs poorly in predictions outside the sample because it was trained too closely to the
in-sample data. When the out-of-sample prediction is a significant issue, a restriction on the number of variables and decision nodes is recommended [12].

The random forest approach avoids overfitting without trimming the tree or limiting the number of divisions allowed by constructing many different trees. The results of the trees are averaged to decrease the variance of the forecast. It also uses a random sampling of variables to divide the data at each node. So, the same variables are not available at the nodes of each tree. Result: overfitting the in-sample data is not a concern in most instances “Tiffin[13], 2016”.

5.2. Support Vector Machine (SVM). An independent and identically distributed dataset (iid) is required in classification applications that use a different machine learning technique. For classification, the data point X is entered into an algorithm. Comparative to machine learning approaches that include calculations of probability distributions, it distributes it to one of the many categorization classes. Particularly, in multidimensional fields, discriminatory approaches that are less effective and used when outlines are required use fewer resources. Last chances are required to identify a multidimensional surface equation that optimally distinguishes numerous classes. This discriminating function can predict new occurrence labels with a high degree of confidence. With its convex optimization problems, SVM always gives the same optimal space value as evolutionary algorithms or perceptions, commonly employed in machine learning classification. The initialization and termination requirements for perceptions are quite substantial [14].

5.3. Logistic Regression. When a binary dependent variable is analyzed using logistic regression, it provides interpretable probabilities that range from zero to one in the relevant circumstance. Additionally, according to this hypothesis, the return on explanatory factors falls as likelihoods approach zero or one. With this increase in production, you will see a massive difference between near zero or one end and close to the other. "Rajkumar [12]".

5.4. Naive Bayes. It was created by British scientist Reverend Thomas Bayes using probabilistic and statistical methods. Naive Bayes performs better than expected in many complicated real-world situations. The simpleness of Naive Bayes makes it a typical model in machine learning since all qualities have an equal influence on the final decision. The Naive Bayes method is appealing because of its simplicity, which translates to computing efficiency. Prior, posterior, and class conditional probabilities are the three primary components of the Naive Bayes Classification [15].

5.5. K-Nearest Neighbors. According to KNN, an instance’s label must match that of its KNN instance. It may also be defined as a case-by-case situation. As far as predictability is concerned, KNN is easy to construct and apparent approaches. Assumptions about data distribution are not made by KNN. Cumulative learning is built on examples that do not require any training before making predictions, and these benefits make it easy for anybody to use it effectively. For classification and regression learning tasks, KNN is commonly employed [16].

5.6. Gradient Boosting. Gradient enhancement is a technique for generating a high-quality preview from a variety of low-quality models. A loss function is used to an initial model of the target variable in most cases. After the loss function is applied to the residues of the previous models, a new model will be displayed. A portion of this procedure continues in [17].

5.7. Neural Network. Humans have around 10 billion neurons in their brains, which are connected in a complicated network. All the linked parts work together to produce intelligent behavior. A neuron’s input signals are made up of the output signals from other neurons linked to it. After reaching a specific threshold, the neuron creates a bioelectric signal that propagates across the synaptic connections to neighboring neurons [18].

Artificial neural models aim to replicate the following properties of this network [19, 20]:

(i) There is simultaneous processing of information by neurons because of parallel processing
(ii) The neuron’s dual role is as a memory and a signal processor
(iii) Distributed data representation: knowledge is dispersed throughout the network, neither preset nor constrained
(iv) The network’s capacity to learn from experience

5.8. Autoregressive Integrated Moving Average Model (ARIMA). When it comes to time series, it is a collection of data that has been collected through time. They may represent both stationary and nonstationary time series and provide reliable predictions based on a description of past data of a single variable. In contrast to previous forecasting models, this one does not make any assumptions about historical data patterns. To create ARIMA models, the Box–Jenkins technique follows these steps: (1) model identification, (2) parameter estimation and selection, (3) diagnostic checking (or modal validation), and (4) model’s use [21].

6. Empirical Results

In this section, we describe the main results of our analysis in four steps.

6.1. The First Step Determines the Accuracy Model. After processing the data on Python to obtain the accuracy results for the algorithms used, we get Table 1.

From Table 1, it is clear that the gradient boosting and logistic regression models are the most accurate with a percentage of 87%, followed by the neural network by 86%,
random forest by 83%, Naïve Bayes by 79%, K-nearest neighbors by 77%, and SVM by 74%. So, we will depend on gradient boosting for FDI (the precision and recall indicators of gradient boosting larger than logistic regression) predictions shown in Table 2.

From Table 2, it is clear that the predicted values using gradient boosting are almost identical to the actual values of FDI inflow, indicating the accuracy and high quality of the forecast. Figure 1 illustrates this.

6.2. Second Step Using the ARIMA Model to Predict Independent Variable. In this step, we will use the ARIMA model to predict independent variables from 2020 to 2030. After we get the future values of these variables, we use them in a gradient boosting model to get the future value of FDI inflow to Egypt in the period 2020–2030. Table 3 shows the future value of independent variables.

From Table 3, we can come up with some indicators in the period 2020–2030:

1. Market size indicators: we find an increase in the value of Egypt G.D.P., G.D.P. per capita, and population size
2. Economic stability: we find a slight increase in the unemployment rate and a decrease in the inflation rate
3. Cost of finance and exchange rates: they decrease in the lending interest rate and increase in the exchange rate (Egyptian pound depreciation helps Egypt exports to compete in other markets)
Human capital: we find improvement in the human development index during the period.

Trade openness: the share of trade in the Egyptian G.D.P. was fluctuating, sometimes increasing, and others decreasing.

We can judge the accuracy of the ARIMA model predictions through Table 4.

Table 5 shows that the mean absolute percentage error (MAPE) value is suitable for all variables and the coefficient of determination (R2) also. This indicates the accuracy of the ARIMA model and stationary time series of the variables.

6.3. The Third Step Is Using the Future Value of the Independent Variable to Get the Future Value of FDI. After we obtain Table 3, we use it for predicting Egypt FDI inflow by gradient boosting classifier, and we find stability in the Egypt FDI inflow during the period. We can show it in Table 5.

6.4. Determine the Importance Variable by Measuring Its Effect on FDI Prediction. Feature importance: these methods are most commonly used for prediction; however, examining the feature importance can help determine which of your variables have the most significant impact on these models. Table 6 shows the outcome of this code.

From Table 6, we find that the preliminary determination of Egypt FDI inflow is the human development index by 40.6%, followed by population size by 21.2%, G.D.P. per capita by 11.6%, lending rate by 10.3%, G.D.P. value by 8%, inflation rate by 4.3%, unemployment rate by 2.4%, and exchange rate and trade openness by 0.8%, respectively.
We can summarize the results in the following points:

1. The gradient boosting is the most accurate model for Egypt’s FDI inflow forecasting.
2. ARIMA model provides us accurate results for independent variables’ prediction.
3. During the current decade (2020–2030), the economic indicators in Egypt are stable.
4. The primary determination of Egypt’s FDI inflow is the human development index, population size, G.D.P. per capita, lending rate, and G.D.P. value.

**7. Conclusion**

Decision makers always seek to predict the values of agreed foreign direct investment and its main determinants, to determine its needs for the financing needed for investment in general on the one hand and, on the other hand, to determine the methods of attracting it to the host country. In this paper, the researcher used machine learning algorithms to predict the foreign investment flowing into Egypt and its determinants. After using the most accurate algorithm, it appears that the foreign direct investment flowing into Egypt in the current decade 2020–2030 is, to a large extent, stable.

As for the determinants of foreign direct investment in Egypt, we found that the primary determinant is: human development index, population size, G.D.P. per capita, lending rate, and G.D.P. value. As for the economic indicators of the size of the market, it became clear that the value of the Egyptian gross domestic product, its per capita share, and the size of the population increased during the current decade. As for the indicators of economic stability, it became clear that there was a slight increase in unemployment rates and a significant decrease in inflation rates to some extent.

In the future, we can apply machine learning algorithms to determine the determinants of economic growth for countries and the main channels of economic policies, such as monetary and fiscal policies. We can also get an accurate prediction of trade relations between the countries of the world.

**Data Availability**

The data used to support the findings of the study can be obtained from the corresponding author upon request.

**Conflicts of Interest**

The authors declare no conflicts of interest.

**Acknowledgments**

This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant no. KEP-PhD-48-130-38. The authors, therefore, acknowledge with thanks DSR technical and financial support.

**References**

[1] World Bank, "dataset, world development indicators," https://databank.worldbank.org/source/world-development-indicators.
[2] "He central bank of Egypt," Economic Review, vol. 58, no. 4, 2017.
[3] D Singh, "Interpretable machine-learning approach in estimating FDI inflow: visualization of ML models with LIME and H2O," TalTech Journal of European Studies, vol. 11, no. 1, 2021.
[4] J. H. Dunning, Multinational Enterprises and the Global Economy, Addison-Wesley, Wokingham, 1993.
[5] J. H. Dunning, "The eclectic paradigm as an envelope for economic and business theories of M.N.E. activity," International Business Review, vol. 9, no. 2, 2000.
[6] H. Dunning and S. Lundan, *Multinational Enterprises and the Global Economies*, Edward Elgar publishing limited, Cheltenham, United Kingdom, 2008.

[7] G. Teskata and E. Gasi, *Determinants of Foreign Direct Investment in Ghana*, Overseas Development Institute, University of Ghana, Accra, Ghana, 2000.

[8] Oecd, *Foreign Direct Investment for Development (Maximizing Benefits, Minimizing Costs)*, OECD, Paris, France, 2002.

[9] M. Asiamah, D. Ofori, and J. Afful, “Analysis of the determinants of foreign direct investment in Ghana,” *Journal of Asian Business and Economic Studies*, vol. 26, no. 1, 2018.

[10] A. Hintosova et al., “Determinants of foreign direct investment inflows: a case of the Visegrad countries,” *Journal of International Studies*, vol. 11, no. 2, 2018.

[11] S. Asongu, S. Uduak, and R. Salisu, “Determinants of foreign direct investment in fast-growing economies: evidence from the BRICS and MINT countries,” *Financial Institution*, Springer, vol. 4, p. 26, 2018.

[12] V. Rajkumar, *Predicting Surprises to G.D.P.: A Comparison of Econometric and Machine Learning Techniques*, Cornell University, Ithaca, New York, 2017.

[13] A. Tiffin, *Seeing in the Dark: A Machine-Learning Approach to Nowcasting in Lebanon*. I.M.F. Working Papers, International Monetary Fund, Washington, DC, United States, 2016.

[14] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Springer, Heidelberg, Germany, 2015.

[15] Y. Nugraha, “Na ¨i ve Bayes classifier for journal quartile classification,” *International Journal of Recent Contributions from Engineering Science & IT*, vol. 7, no. 2, p. 91, 2019.

[16] S. Kang, “K-nearest neighbor learning with graph neural networks,” *Mathematics*, vol. 9, p. 830, 2021.

[17] A. Richardson, T. Mulder, and T. Vehbi, *Nowcasting New Zealand G.D.P. Using Machine Learning Algorithms*, The Australian National University, Canberra, Australia, CAMA Working Paper 47/2018, 2018.

[18] C. Gallo, *Encyclopedia of Information Science and Technology*, IGIGlobal, Hershey, PA, 2015.

[19] Abdel-Minaam D. Salama et al., “An adaptive fitness-dependent optimizer for the one-dimensional bin packing problem,” *IEEE Access*, vol. 8, pp. 97959–97974, 2020.

[20] D. S. Abdelminaam and F. H. Ismail, “Coaid-deep: an optimized intelligent framework for automated detecting covid-19 misleading information on Twitter,” *IEEE Access*, vol. 9, pp. 27840–27867, 2021.

[21] A. Coghlan, *A Little Book of R for Time Series*, Readthedocs.Org, http://a-little-book-of-r-for-time-series.readthedocs.org/en/latest/src/timeseries.html, 2010.