Forecasting Algeria's General Industrial Index: ϕ Comprehensive Data-Driven Analysis and Predictive Insights

Siham Bouguern

December 19, 2023

Forecasting Algeria's General Industrial Index: A Comprehensive Data-Driven Analysis and Predictive Insights

Siham Bouguern
Jawsak Group
Algeria
siham.bouguern@jawsak.com

***Abstract*— The manufacturing sector serves as the cornerstone of a nation's economic, productive, and social influence. Given its substantial share of global energy consumption, embracing data-driven strategies represents a pivotal approach to informed decision-making. Machine learning techniques, renowned for their capacity to address intricate challenges, are increasingly favored for energy forecasting. This research introduces and assesses four analytical models - namely, Linear Regression, Exponential Smoothing, ARIMA, and SARIMA - in the context of predicting Algeria's General Industrial Index. In this study, these four models are applied to assess and predict Algeria's general industrial production for the forthcoming decade (2021–2030) by utilizing a 50-years dataset provided by The National Office of Statistics (NOS). The findings of this study reveal that the four forecasting methods provides a unique perspective on the future trajectory of the General Industrial Index in Algeria. Linear Regression is optimistic, Exponential Smoothing suggests stable growth, ARIMA predicts a decline, and SARIMA reinforces the declining trend with seasonality considerations. This paper significantly contributes to the literature by providing a comprehensive analysis of Algeria's general industrial production using four distinct forecasting models. The findings offer valuable insights for policymakers and industry stakeholders, facilitating informed decisions to sustain and enhance manufacturing output in Algeria.**

Keywords*— *manufacturing industries, index of industrial production, data analytics, forecasting, linear regression, exponential smoothing, ARIMA, SARIMA

I. Introduction

We currently reside in the age of data, where the significance of data lies not in its sheer existence, but rather in its potential to inform decision-making processes. To address this, the field of "Data Analytics" has emerged, drawing on techniques and methods from data mining and statistical analysis, with contributions from scholars and practitioners in artificial intelligence (AI), algorithms, and database domains. These methods are designed to extract actionable insights from vast and diverse datasets [1]. A report by the McKinsey Global Institute in 2011 emphasizes that the analysis of extensive datasets will become a fundamental driver of competitiveness, productivity growth, and innovation, showcasing the diverse applications of big data in value creation [2]. Consequently, businesses are increasingly embracing precision marketing strategies to remain competitive and safeguard their profit margins. As a result, forecasting models have gained widespread use in precision marketing to comprehend and meet customer demands [3].

In the realm of intelligent manufacturing, industrial big data serves a dual purpose. It not only empowers enterprises to accurately perceive changes in the internal and external environment but also facilitates data-driven analysis and decision-making to optimize production processes, reduce costs, and enhance operational efficiency [4]. Predictive analytics, a domain of growing significance, involves the application of statistical techniques to analyze historical data for predicting future events and behaviors [5]. Accurate and sophisticated demand forecasts are pivotal in making informed decisions related to inventory management, purchasing, and assortment planning, especially in the era of Industry 4.0. This era is characterized by reduced fulfillment times and the generation of vast amounts of data from advanced technologies, offering opportunities for data-driven decision-making, particularly in the context of demand forecasting [6]. Classical time series forecasting models have a long-standing history in forecasting within the business environment. Notably, the AutoRegressive Integrated Moving Average (ARIMA) model, known for its accuracy, expresses the time series structure through a relational equation, making its results highly persuasive for decision-makers. The Seasonal AutoRegressive Integrated Moving Average (SARIMA) model extends the capabilities of ARIMA by incorporating seasonality [7].

Global energy consumption is anticipated to surge by more than 50% by the year 2030 [8]. Algeria, in particular, presents an economic landscape marked by its heavy reliance on oil exports, comprising more than 97% of its total exports, and standing as one of the world's foremost exporters of natural gas. Moreover, the Algerian economy exhibits limited diversification in its production base, with a predominant emphasis on gas extraction industries [9]. On one hand, the escalation in gas and oil production is driven by the imperative to satisfy the burgeoning energy demands of households and domestic industries, often characterized by their high energy consumption, a common trait in oil-rich nations. Conversely, on the other hand, the amplification of gas and oil production serves the purpose of bolstering export volumes, thereby securing additional financial resource [9].

In this paper we analyzed the national industrial production in Algeria over the period 1970 - 2020. Using and comparing between the following models: Linear Regression, Exponential Smoothing, ARIMA, and SARIMA. The remainder of the paper is organized as follows: Section 2 represents the state of the art. Section 3 describes the evolution of industrial production in Algeria. Section 4 presents the methodology. Section 5 discusses the critical analysis and findings and section 6 concludes the paper along with some perspectives for future work.

II. State of the Art

A. Data Analysis for Forecasting in Manufacturing Sector

The industrial and manufacturing sector uses the most energy worldwide. To reduce global energy consumption, it's crucial to adopt strategies and techniques that improve energy efficiency and management. Belhadi et al. in [1] have devised an original model for encapsulating key features of Big Data Analytics (BDA) within the domain of manufacturing processes. This was achieved through a synthesis of current research outcomes and a comprehensive case study within a prominent manufacturer of phosphates derivatives. The purpose was to highlight the capabilities of BDA in manufacturing processes and provide suggestions for advancing research in this area. The insights derived from their study are expected to assist companies in comprehending the capabilities of BDA, recognizing potential impacts on their manufacturing processes, and facilitating the development of more efficient BDA-enabling infrastructure, thereby minimizing the risk of plagiarism detection [1]. While Qin & Chiang [2] delineate the trajectory of progress in machine learning and AI, propelled by advancements in statistical learning theory spanning the past two decades and underscored by the commercial achievements of prominent big data enterprises. Subsequently, they delve into an examination of the distinctive attributes of process manufacturing systems, followed by a concise overview of research and development in data analytics over the preceding three decades [2]. The investigation conducted by Wang et al. [4] offers an all-encompassing examination of interconnected subjects, including the definition of big data, as well as methodologies grounded in both model-driven and data-driven approaches. The study delves into the framework, evolution, pivotal technologies, and applications of BDA within the context of intelligent manufacturing systems. Additionally, the exploration accentuates the challenges faced and opportunities presented for prospective research in this domain [4].

The increasing popularity of machine learning techniques for predicting energy consumption is attributed to their proficiency in addressing complex non-linear challenges. Mawson & Hughes [8] undertake a comparative analysis, evaluating the efficacy of two deep neural networks—feed-forward and recurrent—in predicting energy consumption and workshop conditions within manufacturing facilities. The predictive models rely on a combination of production schedules, climatic conditions, thermal properties of the facility building, and considerations of building behavior and use [8]. Amidst the COVID-19 pandemic, numerous studies have been undertaken, and in [5], Sheng et al. offer a review focusing on methodological innovations in the exploration of BDA. The study elucidates on how these innovations can be optimally employed to scrutinize current organizational challenges. Notably, the authors provide insights into methodologies encompassing descriptive/diagnostic, predictive, and prescriptive analytics, elucidating their potential application in the examination of unforeseen and impactful events, exemplified by the global crisis resulting from COVID-19. The study emphasizes the implications of such events for managers and policymakers [5].

Furthermore, the utilization of Predictive Analytics (PA) in the oversight of supply chains has gained considerable prominence in recent years, particularly with a focus on demand forecasting. As expounded in [2], Falatouri et al. offer a comprehensive overview of methodologies in Retail Supply Chain Management (SCM) and undertake a comparative analysis of two selected methods. The analysis is grounded in the examination of more than 37 months of authentic retail sales data obtained from an Austrian retailer. The study involves the training and evaluation of SARIMA and LSTM (Long Short-Term Memory) models, both of which demonstrated satisfactory to commendable results based on the provided data [6]. Seyedan and Mafakheri [3] conduct an inquiry into the applications of predictive BDA in demand forecasting within supply chains. Their investigation aims to categorize these applications, discern gaps in existing research, and offer insights for prospective studies. The authors classify algorithms and their applications in SCM into distinct categories, encompassing time-series forecasting, clustering, KNN (K-nearest-neighbors), neural networks, regression analysis, support vector machines, and support vector regression [3]. The advancement of AI has facilitated remarkably precise demand forecasting. However, it is essential to recognize that achieving heightened forecast accuracy does not automatically result in reduced inventory costs or enhanced service levels within supply chain and inventory management. In response to this consideration, Shibayama et al. [7] have devised a framework for demand forecasting that prioritizes both accuracy and interpretability. The framework leverages time series decomposition and ARIMA methods. Specifically, Seasonal-trend decomposition using locally estimated scatterplot smoothing (STL) is employed to break down a time series into its components—trend, seasonality, and residual—providing decision makers with a comprehensible foundation for understanding changes in demand. This approach is designed to support decision-makers in making informed choices in the realm of demand forecasting [7].

B. CO2 Emissions Forecasting

The article by Dragomir et al. [10] centers on the evaluation of CO2 emissions and their correlation with the national industrial production of Romania. Through an examination of CO2 emissions and the production of primary industrial goods, the study draws two noteworthy conclusions: Firstly, the trajectory of total production significantly influences emissions. Secondly, alterations in the industrial structure contribute distinctly to emissions during specific periods, such as 1993–1998 and 2002–2012, wherein increased production aligns with corresponding changes in the CO2 emission trends [10]. In the study by Zhou et al. [11] conducted in China, three primary sources of carbon emissions spanning the period from 1990 to 2017 were identified: the energy industry, fuel combustion in other industries, and industrial processes. The paper outlines the development of a driving force model for each emission source using multiple linear regression. Utilizing these models, the study generates forecasts for both carbon intensity and total CO2 emissions from the year 2018 to 2030. The findings indicate a continued reduction in both CO2 emission intensity and total emissions; however, the study underscores the necessity for additional efforts to attain the objectives outlined in the Paris Agreement [11].

Within the realm of developing nations, Algeria emerges as a noteworthy contributor to CO2 emissions, securing the third position among African countries in this context. Addressing this concern, Bouziane et al. [12] conducted similar studies in Algeria. In their work, they delineate a hybrid approach employing Artificial Neural Networks

(ANN) and an agent-based architecture for predicting carbon dioxide (CO2) emissions originating from various energy sources in the city of Annaba, utilizing authentic data. The system is composed of multiple autonomous agents, classified into two types: forecasting agents, responsible for predicting the production of specific energy types using ANN models, and core agents, tasked with critical functions such as calculating equivalent CO2 emissions and managing the simulation [12]. In their analytical review, Hicham & Moussa [13] investigate the interconnections among energy consumption, economic growth, and CO2 emissions in Algeria spanning the period from 1970 to 2017. Employing a structural Vector Autoregression (VAR) approach, the study's outcomes indicate that a positive shock in CO2 emissions corresponds to an increase in both economic growth and energy consumption. Additionally, the study reveals that a positive shock in energy consumption has a minimal positive effect on economic growth but a substantial negative impact on CO2 emissions [13]. The paper by Bouznit and Pablo-Romero [9] seeks to examine the relationship between CO2 emissions and economic growth in Algeria, considering factors such as energy use, electricity consumption, exports, and imports. The study assesses the validity of the Environmental Kuznets Curve (EKC) hypothesis for the period spanning 1970 to 2010, employing the Autoregressive Distributed Lag model extended to incorporate break points. The findings affirm the applicability of the EKC hypothesis for Algeria, suggesting a nonlinear relationship between CO2 emissions and economic growth during the specified time frame [9]. The study of Chekouri et al. [14] conducted a stochastic impact by regression on population, affluence, and technology model to identify the determinant factors driving CO2 emissions in Algeria during the period 1971–2016. The method of partial least squares regression is applied to eliminate multicollinearity problems. The results indicate that the population has a positive and significant effect on CO2 emission. The energy use is found to be the second most contributing factor to CO2 emissions followed by urbanisation and affluence (GDP per capita) [14]. In their work, Abounoori and Bagherpour [15] introduced a Hybrid Regression Neural Network approach with the aim of achieving improved fitness compared to traditional Regression Analysis and Neural Network methods. Through a comparative analysis of the estimated results obtained from Regression Analysis and Neural Networks against those derived from the Hybrid Neural-Regression method, the study highlights the superior performance of the latter approach [15].

III. Analysis of the Evolution of Industrial Production in Algeria (1970 - 2020)

In Algeria, the industrial sector witnessed significant growth in the 1970s and early 1980s. However, during the 1990s, this sector, and the overall economy, faced a period of weak performance. The Ministry of Industry and Restructuring reported a 25.8% decline in industrial production between 1989 and 1998, with production capacity utilization remaining low at around 30% to 60%. Additionally, the industrial sector's value-added saw insufficient representation, and exports of industrial products stagnated.

However, in 1998, there was a notable recovery, as indicated by a 10.50% growth in the industrial production index compared to previous years, according to ministry data. The National Office of Statistics (NOS) in 2010 attributed the declining growth rate to a significant decrease in hydrocarbon production levels in the fourth quarter of 2008 (-7.5%) and the first three quarters of 2009 (-9.9%, -8.7%, and -3.6%), resulting in an overall growth rate of -5.4% for the year.

Despite the overall decline in industrial activity from 1998 to 2008, dropping from 69.0% to 52.8%, manufacturing industries managed to achieve a modest growth of 1.9% in 2008, even though there was a 2.4% decline in the third quarter of the same year following several years of negative growth rates since 1999. In more recent years, the industrial sector in Algeria has shown some resilience and growth, as indicated by the data for 2010-2020, though with fluctuations [16]. Figure 1 presents the evolution and the variations of the annual index of industrial production of the national public sector:

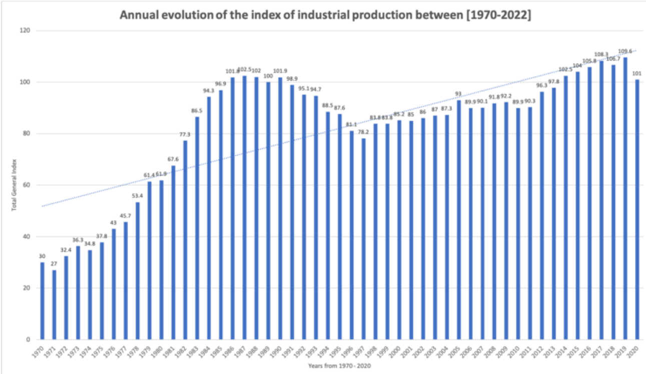

Fig. 1. Annual evolution of the index of industrial production between [1970-2020]

IV. Methodology

A. Data Collection and Preprocessing

The datasets of the General Index were collected by The National Office of Statistics (NOS) of Algeria [17] starting from 1970 until 2020. The 50 years of values recorded for all industries including: Water-Energy, Hydrocarbons Industries, Mines and Quarries, S.M.M.E.E.I (Steel, Metallurgical, Mechanical, Electrical and Electronic Industries), Materials of Construction, Ceramic, Glass, Chemistry, Rubber, Plastics, Agro-Food industry, Tobacco, Matches, Textile industries, Leather and Footwear Industries, Wood, Cork, Paper Industries, and other diverse Industries. In addition to the Total General Index, Total non-hydrocarbon industries index, and Total Manufacturing Industries Index.

```
# Check for missing values
data.isnull().sum()

Year    0
GI      0
dtype: int64
```

Fig. 2. Python code snippet for checking missing values

Figure 2 shows the results of missing value search, as indicated there are no missing values.

B. Descriptive Analysis of the Used Data

In Figure 3 the main information of the variables used in this analysis are shown. While Figure 4 presents the main descriptive statistics of the dataset used in this analysis.

```
Information on Dataframe:
=========================
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Year    51 non-null     int64
 1   GI      51 non-null     float64
dtypes: float64(1), int64(1)
memory usage: 944.0 bytes
None
```

Fig. 3. Python code output snippet for dataset information

Descriptive Statistics:
==

	count	mean	std	min	25%	50%	75%	max
Year	51.0	1995.000000	14.866069	1970.0	1982.50	1995.0	2007.50	2020.0
GI	51.0	82.076471	23.337134	27.0	77.75	89.9	98.35	109.6

Fig. 4. Python code output snippet for descriptive statistics

To display the general trend of the Algerian Total General Index from 1970 to 2020, Figure 5 is presented.

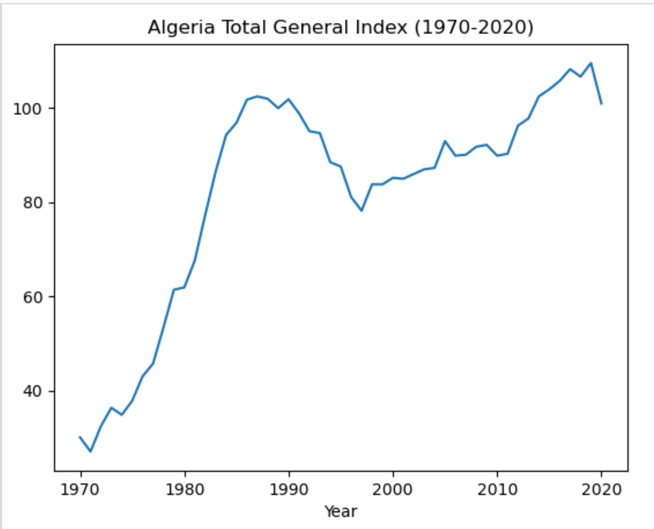

Fig. 5. Algeria total general index [1970-2020]

C. *Selection of Forecasting Methodologies*

In selecting a forecasting model, careful consideration of the data's characteristics, model interpretability, and underlying assumptions is crucial. With the available models, including Linear Regression, Exponential Smoothing, ARIMA, and SARIMA, the choice primarily depends on the specific context. Here are the main selection criteria:

- Data Characteristics: Considering the nature of time series data, including its stationarity, seasonality, and trends.
- Interpretability: Evaluating the model's ability to provide interpretable forecasts and insights.
- Simplicity: Choosing a model that balances complexity with the research's practicality.
- Robustness: Assessing the model's resistance to outliers and extreme values in the data.
- Relevance: Ensuring the selected model aligns with the research objectives and domain knowledge.
- Performance Validation: Conducting rigorous performance assessment, such as cross-validation, to gauge the model's accuracy.
- Limitations: Acknowledging and consider the limitations inherent to the chosen model.
- Seasonality: If present, select models capable of accounting for seasonality patterns.
- Domain Expertise: Leveraging domain-specific knowledge when making the final model selection.

V. Critical Analysis and Findings

A. *Method 1: Linear Regression*

The Linear Regression model suggests a relatively optimistic outlook for the General Industrial Index in Algeria. The forecast indicates a consistent upward trend in the Index over the next ten years.

```
Forecast results Using Linear Regression:
=========================================
   Year  Forecasted Index
0  2021        113.510118
1  2022        114.719104
2  2023        115.928090
3  2024        117.137077
4  2025        118.346063
5  2026        119.555050
6  2027        120.764036
7  2028        121.973023
8  2029        123.182009
9  2030        124.390995
=========================================
```

Fig. 6. Forecast results of the Linear Regression Model

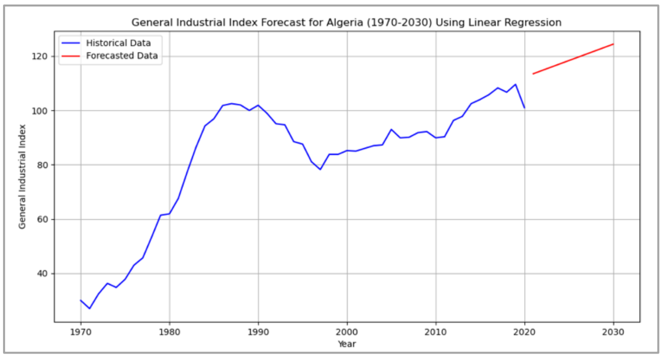

Fig. 7. General industrial index forecast for Algeria using Linear Regression Model

According to this model, the industrial sector is expected to experience a steady and positive growth trajectory, with the Index increasing at an annual rate of approximately 1.22 index points. This implies confidence in the Algerian industrial sector's potential for expansion and development (See Figure 7).

B. *Method 2: Exponential Smoothing*

The Exponential Smoothing model provides a forecast with a relatively stable trajectory for the General Industrial Index in Algeria. It anticipates a moderate and consistent growth pattern with minor fluctuations.

Forecast results Using Exponential Smoothing:

	Year	Forecasted Index
0	2021	100.505293
1	2022	100.685924
2	2023	100.216032
3	2024	101.746320
4	2025	100.751280
5	2026	103.081338
6	2027	103.461049
7	2028	101.666272
8	2029	103.546164
9	2030	101.751232

Fig. 8. Forecast results of Exponential Smoothing Model

Although the growth rate is relatively low, this model suggests that the Index will maintain its stability over the coming decade, with occasional short-term variations. It portrays a scenario of steady, albeit slow, progress in the industrial sector (See Figure 9).

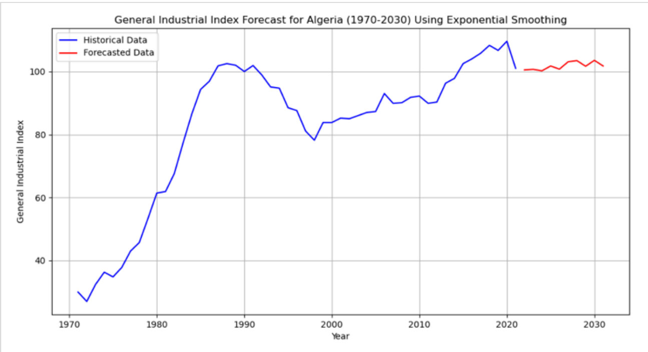

Fig. 9. General industrial index forecast for Algeria using Exponential Smoothing Model

C. *Method 3: ARIMA*

The ARIMA model, in contrast to the previous models, paints a less optimistic picture for the General Industrial Index in Algeria. It forecasts a declining trend in the Index over the next ten years, indicating potential challenges or structural changes within the industrial sector.

The forecasted decrease in the Index is relatively significant, with the Index declining each year. This model implies that the industrial sector might face headwinds and contraction in the near future (See Figure 11).

D. *Method 4: SARIMA*

The SARIMA model, which accounts for seasonality and trends in the data, presents a forecast that aligns with the declining trend seen in the ARIMA model. It suggests a decreasing trajectory for the General Industrial Index in Algeria over the next ten years.

Forecast results Using ARIMA:

	Year	Forecasted Index
0	2021	99.232382
1	2022	97.963523
2	2023	94.459775
3	2024	93.549295
4	2025	92.076708
5	2026	90.539457
6	2027	89.808227
7	2028	88.831117
8	2029	88.069560
9	2030	87.531452

Fig. 10. Forecast results of ARIMA Model

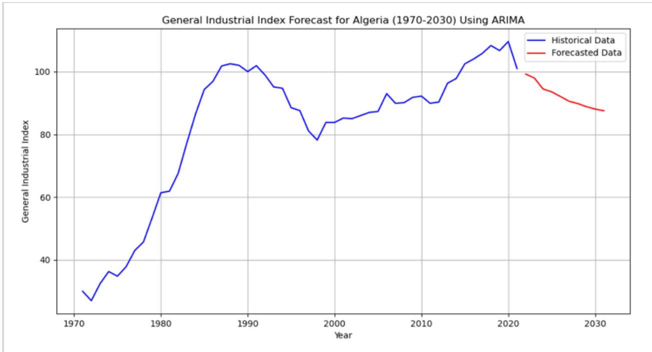

Fig. 11. General industrial index forecast for Algeria using AMIRA Model

While the decline is not as sharp as in the ARIMA model, it still indicates a challenging period ahead for the industrial sector. The SARIMA model highlights the importance of considering seasonality in the context of the forecast (See Figure 13).

Forecast results Using SARIMA:

	Year	Forecasted Index
0	2021	98.152887
1	2022	98.029634
2	2023	96.769332
3	2024	98.208912
4	2025	97.305021
5	2026	99.395185
6	2027	99.709836
7	2028	99.063797
8	2029	101.991889
9	2030	99.660582

Fig. 12. Forecast results of SARIMA Model

In summary, each of the four forecasting methods provides a unique perspective on the future trajectory of the General Industrial Index in Algeria. Linear Regression is optimistic, Exponential Smoothing suggests stable growth, ARIMA predicts a decline, and SARIMA reinforces the declining trend with seasonality considerations.

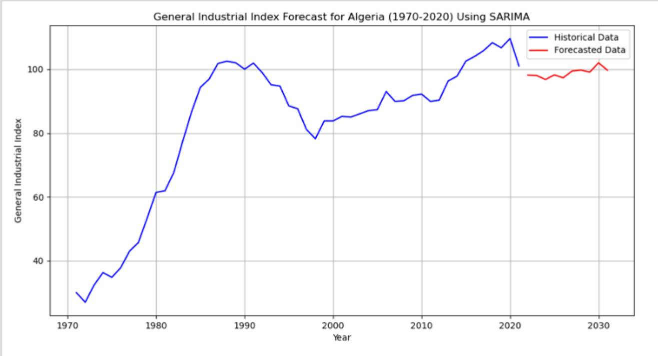

Fig. 13. General industrial index forecast for Algeria using SAMIRA Model

VI. Conclusion and Future Work

In this paper we applied four different methods to test and forecast the general industrial production for Algeria over the next 10 years (2021–2030). Using the 30 years dataset collected by The National Office of Statistics (NOS), we observed the trend of Algerian production of main industrial. In general, there is a degree of disagreement among the models regarding the future trend of the General Industrial Index in Algeria. Linear Regression forecasts steady growth, Exponential Smoothing suggests stability with minor fluctuations, ARIMA predicts a declining trend, and SARIMA indicates a decreasing trend with some variations.

Our major contribution to the literature is the analysis of industrial production. By collecting and evaluating this data, technical measures can be taken to maintain a good level of manufacturing production in Algeria. In conclusion, forthcoming studies could expand these results with additional appraisal of the different common factors. This assessment of the uncertainties in a resolution is an essential element of reasonable, accurate information policy-making. Future work could study and select the most useful method given the domain knowledge.

References

[1] A. Belhadi, K. Zkik, A. Cherrafi, Y. M. Sha'ri and S. E. fezazi, "Understanding the capabilities of Big Data Analytics for manufacturing process: insights from literature review and multiple case study," Computers & Industrial Engineering, 2019.

[2] S. J. Qin and L. H. Chiang, "Advances and opportunities in machine learning for process data analytics," Computers and Chemical Engineering, 2019.

[3] M. Seyedan and F. Mafakheri, "Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities," Journal of Big Data, 2020.

[4] J. Wang, C. Xu, J. Zhang and R. Zhong, "Big data analytics for intelligent manufacturing systems: A review," Journal of Manufacturing Systems, 2021.

[5] J. Sheng, J. Amankwah-Amoah, Z. Khan and X. Wang, "COVID-19 Pandemic in the New Era of Big Data Analytics: Methodological Innovations and Future Research Directions," British Journal of Management, 2021.

[6] T. Falatouri, F. Darbanian, P. Brandtner and C. Udokwu, "Predictive Analytics for Demand Forecasting – A Comparison of SARIMA and LSTM in Retail SCM," in 3rd International Conference on Industry 4.0 and Smart Manufacturing, 2022.

[7] H. Shibayama, A. Ishigaki, T. Kobe, T. Ueda, D. Arimizu and T. Nakamura, "Design of a Framework for Demand Forecasting Using Time Series Decomposition-Based Approach," in 4th Asia Pacific Conference of the Prognostics and Health Management, 2023.

[8] V. J. Mawson and B. R. Hughes, "Deep learning techniques for energy forecasting and condition monitoring in the manufacturing sector," Energy & Buildings, 2020.

[9] M. Bouznit and M. d. P. Pablo-Romero, "CO2 emission and economic growth in Algeria," Energy Policy, p. 93–104, 2016.

[10] C. M. B. Dragomir, G. Podaru and C. Muntenita, "A Bayesian Analysis of CO2 Emissions and National Industrial Production in Romania," in Mechanisms and Machine Science, 2022, pp. 683-690.

[11] Y. Zhou, J. Zhang and S. Hu, "Regression analysis and driving force model building of CO2 emissions in China," Scientific Reports, p. 2021.

[12] S. E. Bouziane, M. T. Khadir and J. Dugdale, "A collaborative predictive multi-agent system for forecasting carbon emissions related to energy consumption," Multiagent and Grid Systems, pp. 39-58, 2021.

[13] A. Hicham and C. Moussa, "A Structural VAR analysis of Energy Consumption, Economic Growth and CO2 Emissions: Evidence from Algeria," Journal of Managerial and Economic Research, pp. 101-114, 2019.

[14] S. M. Chekouri, A. Chibi and M. Benbouziane, "Examining the driving factors of CO2 emissions using the STIRPAT model: the case of Algeria," International Journal of Sustainable Energy, 2020.

[15] E. Abounoori and M. Bagherpour, "Estimation of Industrial Production Costs, Using Regression Analysis, Neural Networks or Hybrid Neural - Regression Method?," Iranian Economic Review, vol. 11, no. 16, pp. 17-29, 2006.

[16] N. Bouri, "The manufacturing sector in Algeria: analysis and interpretation of annual variations," Revue du Lareiid, 2020.

[17] "Industrial Production Index - National Public Sector: 1970 to 1976," The National Office of Statistics, 2020.

Appendices

Appendix A: Code snippet

1.1 Importing Libraries & Dataset

```python
# Import necessary libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

from statsmodels.tsa.holtwinters import ExponentialSmoothing
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.statespace.sarimax import SARIMAX

from datetime import datetime, timedelta

import warnings
from warnings import catch_warnings
from warnings import filterwarnings
warnings.filterwarnings("ignore")
```

```python
# Create dataset
data = pd.read_csv('Index of Industrial Production.csv')
```

1.2 Exploratory Data Analysis (EDA)

```python
# Display the first few rows of the dataset to inspect the data
data.head()
```

	Year	GI
0	1970	30.0
1	1971	27.0
2	1972	32.4
3	1973	36.3
4	1974	34.8

```python
# Extract the year and index data
years = data['Year'].values.reshape(-1, 1)
index = data['GI'].values
```

```python
# Check for missing values
data.isnull().sum()
```

```
Year    0
GI      0
dtype: int64
```

2 Method 1: Linear Regression

```python
# Create a linear regression model
model = LinearRegression()
model.fit(years, index)
```

```
▾ LinearRegression
LinearRegression()
```

```python
# Predict the index for the next 10 years
future_years = np.arange(2021, 2031).reshape(-1, 1)
forecasted_index = model.predict(future_years)
```

```python
# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 6))
plt.scatter(years, index, label='Historical Data', color='b')
plt.plot(future_years, forecasted_index, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using Linear Regression')
plt.legend()
plt.grid(True)

# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using Linear Regression.png')

# Show the plot
plt.show()
```

3 Method 2: Exponential Smoothing

```
# Extract the year and index data
years = data['Year']
index = data['GI']
# Create a DataFrame with the extracted data
df = pd.DataFrame({'Year': years, 'Index': index})
```

```
# Generate a date-based index
start_year = df['Year'].min()
end_year = df['Year'].max()
date_index = pd.date_range(start=f"{start_year}-01-01", end=f"{end_year}-12-31", freq='A')
# Set the date-based index for the DataFrame
df.set_index(date_index, inplace=True)
```

```
# Perform exponential smoothing forecasting
model = ExponentialSmoothing(df['Index'], seasonal='add', seasonal_periods=12)
result = model.fit()
```

```
# Forecast the index for the next 10 years
forecasted_index = result.forecast(steps=10)
```

```
# Generate a date-based index for the forecasted values
future_dates = [df.index[-1] + timedelta(days=365*i) for i in range(1, 11)]
forecasted_index.index = future_dates
```

```
# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 6))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using Exponential Smoothing')
plt.legend()
plt.grid(True)
# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using Exponential Smoothing.png')
# Show the plot
plt.show()
```

4 Method 3: ARIMA

```
# Fit an ARIMA model
model = ARIMA(df['Index'], order=(5, 1, 0))  # You can adjust the order parameter based on your data
result = model.fit()
```

```
# Forecast the index for the next 10 years
forecasted_index = result.predict(start=len(df), end=len(df) + 9)
```

```
# Generate a date-based index for the forecasted values
future_dates = [date_index[-1] + pd.DateOffset(years=i) for i in range(1, 11)]
forecasted_index.index = future_dates
```

```
# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 6))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2030) Using ARIMA')
plt.legend()
plt.grid(True)

# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using ARIMA.png')

# Show the plot
plt.show()
```

5 Method 4: SARIMA

```
# Fit a SARIMA model
# You can adjust the order and seasonal_order parameters based on your data and domain knowledge
order = (1, 1, 1)
seasonal_order = (1, 1, 1, 12)
model = SARIMAX(df['Index'], order=order, seasonal_order=seasonal_order)
result = model.fit()

# Forecast the index for the next 10 years
forecasted_index = result.get_forecast(steps=10)

# Generate a date-based index for the forecasted values
future_dates = [date_index[-1] + pd.DateOffset(years=i) for i in range(1, 11)]
forecasted_index = forecasted_index.predicted_mean
forecasted_index.index = future_dates

# Plot the historical data and the forecasted values
plt.figure(figsize=(12, 6))
plt.plot(df.index, df['Index'], label='Historical Data', color='b')
plt.plot(forecasted_index.index, forecasted_index.values, label='Forecasted Data', color='r')
plt.xlabel('Year')
plt.ylabel('General Industrial Index')
plt.title('General Industrial Index Forecast for Algeria (1970-2020) Using SARIMA')
plt.legend()
plt.grid(True)

# Save the figure
plt.savefig('General Industrial Index Forecast for Algeria (1970-2030) Using SARIMA.png')

# Show the plot
plt.show()
```