Table of Contents

Table of Tables

Table of Figures

Abstract

Since the 1930s, when oil was first found, the Kingdom of Saudi Arabia's industrial sector has significantly contributed to the country's economic growth. The objective of this project is to create a prediction model for the industrial production index (IPI) in the kingdom in order to meet the objectives specified in Vision 2030, which is focused on economic diversification. The project's goal is to address the difficulties involved in precisely predicting and measuring industrial production, including seasonal and cyclical patterns, outside factors and volatility, and the requirement for sector-specific analyses. The research attempts to anticipate the IPI for the following six months using statistics from the General Authority for Statistics (GASTAT) and auto-regressive integrated moving average (ARIMA) models and decision tree model. In order to promote decision-making, resource allocation, and long-term growth in Saudi Arabia's industrial sector, the study underlines the significance of trustworthy forecasting models and analytical frameworks. This research helps to clarify the trajectory and prospects of the industrial sector by examining historical trends and IPI behavior. As a result, policymakers, economists, and business professionals may make more informed decisions.

1 Introduction

The industrial history of the Kingdom of Saudi Arabia spans a considerable period. Previously, the industrial sector in Saudi Arabia was predominantly focused on handicrafts. However, the advent of the oil discovery in 1930 marked the true onset of industrialization in the country.

To fulfill the objectives of Vision 2030, the Kingdom of Saudi Arabia initiated the development of its economy through the industrial sector. The industrial sector plays a significant and crucial role in the kingdom. The industrial production index is a comprehensive metric that gauges short-term variations in production volume within a specified period [1] [2].

Therefore, to accomplish the aims outlined in Vision 2030, we will examine the seasonal and trend effects on this indicator as well as forecast its growth in upcoming periods. The base year employed in our project is 2010.

This project involves an exploration of the behavior of the time series pertaining to the industrial production index in the Kingdom using auto-regressive integrated moving average (ARIMA) models by utilizing real-data applications and numerical simulations alongside the development of a predictive model to estimate the index value for the subsequent six months using a combination of datasets from the GASTAT (General Authority for Statistics) Statistical Database.

1.1 Problem Statement

As a result of the complex and dynamic nature of the industrial sector in Saudi Arabia, forecasting and analyzing industrial production accurately poses significant challenges. To effectively capture the unique characteristics and factors influencing industrial production in the country, robust forecasting models and analytical frameworks are needed. Moreover, forecasting and analyzing industrial production accurately is further complicated by the absence of forecasting and analyzing up-to-date data, as well as external factors such as global economic trends and geopolitical events.

Several key issues need to be addressed:

1. Seasonality and Cyclical Patterns: Industrial production in Saudi Arabia is influenced by various seasonal and cyclical patterns. These patterns can be challenging to identify and model accurately,

leading to forecasting errors. Developing robust methodologies that account for seasonality, cyclical fluctuations, and other temporal dependencies is essential for accurate industrial production forecasting and analysis.

2. External Factors and Volatility: Saudi Arabia's industrial production is influenced by external factors such as global economic trends, geopolitical events, and fluctuations in commodity prices. These factors introduce volatility and uncertainty into the forecasting process. Accounting for these external factors and their potential impact on industrial production is crucial for accurate analysis and forecasting.

3. Sector-specific Analysis: The industrial sector in Saudi Arabia comprises various sub-sectors, each with its unique characteristics and dynamics. Analyzing and forecasting industrial production at a granular level requires sector-specific analysis and models that capture the intricacies of each sub-sector. Developing tailored analytical frameworks for different industrial sectors is essential for accurate forecasting and analysis.

We will evaluate the data and determine how these challenges are critical for politicians, economists, and industry professionals to make educated decisions, maximize resource allocation, and support long-term growth in Saudi Arabia's industrial sector. Stakeholders may effectively examine historical trends, identify growth possibilities, and reduce risks by establishing reliable forecasting models and analytical frameworks, thereby supporting the overall economic development and diversification goals of the country [3].

1.2 Background

1.2.1 Saudi industrial sector evolution

The Industrial Production Index (IPI) in Saudi Arabia serves as a crucial economic indicator that measures the level of industrial activity and output within the country. It provides valuable insights into the performance and growth of the industrial sector, which plays a significant role in the Kingdom's economy. The history of industrialization in Saudi Arabia can be traced back to the early 20th century, but it gained substantial momentum following the discovery of oil in 1930. The abundant oil reserves in the country prompted the development of various industries, including petrochemicals, refining, manufacturing, and mining. The industrial sector has since become a vital pillar of the Saudi Arabian economy, contributing significantly to employment, GDP, and export revenues.

However, the IPI is designed to capture the short-term changes in the volume of industrial production over a specified period, relative to a selected base period. It provides a quantitative measure of the industrial sector's performance, reflecting the expansion or contraction of output and the overall economic activity associated with industrial production. The index considers various factors such as manufacturing, mining, and electricity generation to provide a comprehensive representation of industrial output.

Moreover, the Saudi Arabian government recognizes the importance of the industrial sector in driving economic growth and diversification. In line with this, Vision 2030, introduced in April 2016 by Crown Prince Mohammed bin Salman, aims to enhance, and expand the country's economic capacity, including the development of the industrial sector. Vision 2030 sets ambitious goals for increasing the contribution of non-oil sectors, including industry, to the nation's GDP.

The Kingdom of Saudi Arabia has witnessed significant development and growth in its industrial sector over the past four decades. The government has implemented policies and initiatives to attract foreign investment, promote local industries, encourage innovation, and foster technological advancements. These efforts have resulted in the establishment of major industrial entities such as Saudi Aramco and Sabic, which have played a pivotal role in driving industrial production and economic progress.

Analyzing and interpreting the IPI provides policymakers, economists, and investors with valuable insights into the performance and trends of the industrial sector in Saudi Arabia. It helps in monitoring the overall economic health, identifying growth opportunities, assessing the effectiveness of policy measures, and making informed decisions regarding resource allocation, investment strategies, and industrial planning.

To assess and monitor the performance of the industrial sector, the industrial production index serves as a crucial indicator. In our project, the study year for the analysis is 2010–2023.

Our project aims to delve into the time series analysis of the industrial production index in the Kingdom of Saudi Arabia, examining its patterns, trends, and potential implications.

Additionally, to anticipate the value of the index for the upcoming six-month period, contributing to a deeper understanding of the industrial sector's trajectory and prospects, we will present it visually using statical graphs. To do so, we will primarily be using the following variables to derive our results:

1. IPI (Industrial Production Index) or (General IPI):

The Industrial Production Index (IPI) is a key economic indicator that measures the level of industrial activity and output within the country It provides valuable insights into the performance and growth of the industrial sector, which plays a significant role in the Kingdom's economy. The IPI considers multiple sectors, including manufacturing, mining, and electricity generation, to present a comprehensive picture of industrial production. By monitoring the IPI, policymakers, economists, and investors can evaluate the overall health of the industrial sector and make informed decisions regarding economic planning, resource allocation, and policy formulation.

2. Mining and Quarrying (IIPQ):

The Mining and Quarrying Index (IIPQ) in Saudi Arabia is designed to measure the level of output and activity within the mining and quarrying sector. It focuses specifically on industries involved in the extraction of natural resources such as oil, gas, minerals, and metals. The IIPQ provides insights into the performance and trends of the mining sector, which is a significant contributor to the Saudi Arabian economy. It helps policymakers and industry experts in monitoring the production levels, identifying opportunities for resource exploration and development, and making informed decisions regarding investment and policy measures in the mining and quarrying sector.

3. Manufacturing (IIPM):

The Manufacturing Index (IIPM) specifically focuses on measuring the performance and output of the manufacturing sector. It provides a detailed analysis of industrial production within the manufacturing industry, which encompasses various sub-sectors such as textiles, chemicals, machinery, and automotive. The IIPM is a critical tool for assessing the growth, trends, and fluctuations in manufacturing activity. It aids policymakers in formulating strategies to enhance the competitiveness of the manufacturing sector, promote technological advancements, and stimulate job creation and export revenues.

4. Electricity and Gas (IIPE):

The Electricity and Gas Index (IIPE) measures the level of output and activity within the electricity and gas sector. This index specifically focuses on the production and generation of electricity, as well as the extraction and distribution of natural gas. The IIPE provides insights into the dynamics of the energy sector, which is vital for industrial production and overall economic growth. By monitoring changes in the IIPE, policymakers and energy experts can assess the capacity, efficiency, and reliability of the electricity and gas infrastructure. It helps in planning and optimizing energy resources, ensuring a stable and sustainable supply for industrial and domestic consumption.

Overall, the Industrial Production Index in Saudi Arabia serves as a key indicator for measuring the level of industrial activity and output. It reflects the growth and performance of the industrial sector, which is a crucial driver of the Kingdom's economy. Understanding and analyzing the IPI enables stakeholders to assess the industrial sector's contribution, identify trends, and make informed decisions to promote sustainable economic growth and development in Saudi Arabia.

2 Literature review

2.1 Modeling and Forecasting Industrial Electricity Demand for Saudi Arabia: Uncovering Regional Characteristics

The production of goods and services in modern economies heavily relies on energy, making it a crucial component. Due to factors such as energy sector deregulation, high price volatility of energy products, and policies related to climate change and energy security, research interest in industrial electricity demand modeling has increased. Understanding the income and price elasticities of industrial electricity demand could aid in formulating effective policies on energy and environmental issues. The literature has used various approaches and dimensions for the econometric analysis of industrial energy demand modeling, including time-varying and fixed coefficient specifications, asymmetric price responses (ARP), and technical changes. Previous studies have used a single equation approach with fixed coefficients and assumed symmetric price elasticity, but recent literature has criticized this assumption as estimates of constant elasticity fail to capture how elasticities vary over time. Moreover, the relationship between industrial electricity demand, income, and prices may not remain constant over time. To address the non-linearity in the price and income elasticities, recent studies have used the time-variable coefficient (TVC) approach, which can detect structural changes and outliers with parameter drift and auxiliary residuals. Some studies argue that technical progress can influence energy demand exogenously, which can be captured by a stochastic trend that can account for the effect of exogenous technological change, the persistence of habits, changes in economic structure, and changes in building and environmental regulation.

The study highlights the importance of energy as a crucial component in the production of goods and services in modern economies, and the increasing research interest in industrial electricity demand modeling due to factors such as energy sector deregulation, high price volatility of energy products, and policies related to climate change and energy security. The income and price elasticities of industrial electricity demand are essential for formulating effective policies on energy and environmental issues. The literature has used various approaches and dimensions for the econometric analysis of industrial energy demand modeling, including time-varying and fixed coefficient specifications, asymmetric price responses (ARP), and technical changes.

Recent studies have used the time-variable coefficient (TVC) approach to address the non-linearity in the price and income elasticities, which can detect structural changes and outliers with parameter drift and auxiliary residuals. Some studies argue that technical progress can influence energy demand exogenously, which can be captured by a stochastic trend that can account for the effect of exogenous technological change, the persistence of habits, changes in economic structure, and changes in building and environmental regulation. In the context of Saudi Arabia, there is limited literature on industrial electricity demand, and the existing studies overlook the role of endogenous and exogenous technological progress in modeling industrial electricity demand. Additionally, the assumption of constant price elasticity of industrial electricity demand needs to be tested in the case of the Saudi economy because of significant structural changes over time.

This study aims to address these limitations and contribute to the literature on industrial electricity demand modeling in Saudi Arabia [3].

2.2 Modeling Industrial Energy Demand in Saudi Arabia and Understanding Its Drivers

A study focuses on modeling aggregate industrial energy demand in Saudi Arabia and identifying the drivers of its growth. The estimated econometric model reveals a long-run income elasticity of 0.60, suggesting that industrial energy consumption will continue to grow as economic activity expands. However, the long-run price elasticity of -0.34 indicates the potential for mitigating this growth through increased energy prices. Industrial companies are found to be more responsive to changes in energy prices than households.

The study also highlights the potential for substantial increases in industrial energy productivity by shifting away from energy-intensive manufacturing. A 10-percentage point shift away from energy-intensive exports could lead to a 6.7% reduction in industrial energy consumption in the long run, contributing to higher value-added manufacturing. Decomposition analysis is applied to understand the drivers of growth in Saudi Arabia's industrial energy consumption. The results indicate that the primary driver is the activity effect, with energy-intensive manufacturing exerting upward pressure on energy consumption. However, the efficiency effect has helped mitigate some of the growth from 2010 onwards. Energy prices have played a limited role, but the energy price reform program implemented in 2015 is expected to further reduce growth rates of industrial energy consumption.

While the study quantifies the impact of higher energy prices on consumption, it does not explore the potential impact on competitiveness. Higher energy prices may weaken the advantage of Saudi Arabia's energy-intensive exports but could also drive the production and export of higher value-added goods. The government may combine higher energy prices with a subsidy scheme to promote industrial energy efficiency and mitigate the negative impact on competitiveness [4].

2.3 Industrial Development in Saudi Arabia: Disparity in Growth and Development

The growth of Saudi Arabia's industrial sector has long been recognized as a significant factor in the social and economic development of the nation. But several problems and worries have prevented this industry's expansion and development. This essay covers much research that sheds light on the variables influencing Saudi Arabia's industrial production and make recommendations for raising this sector's productivity.

One of the main issues identified in the studies is the underutilization of financial resources in the manufacturing industry in India and Saudi Arabia. Lack of suitable training and labor problems were also identified as factors responsible for lower growth and development in the Indian manufacturing industry. To address these issues, Al Bakr [5] suggests reviewing the finance schemes and streamlining the linkage of small industry with large industry regarding research, development, and manufacturing. recommends launching promotional schemes for the growth and development of MSMEs in Saudi Arabia, while Keyed and Kabir Hassan [6] focus on the development of entrepreneurship skills of Saudi citizens.

Infrastructure also plays a crucial role in the development of manufacturing industries. Mehta and Rajan [7] suggest that world-class infrastructure is necessary for the development of the manufacturing industry in Saudi Arabia. They further add that the connectivity of big cities or industrial cities is essential for the smooth conveyance of raw materials, as well as finished goods. Maghrabi Jafery and Sabban [8] observed that SMEs in Saudi Arabia have the potential for growth and ultimately increase oil revenue and economy. They further added that most of the small manufacturing companies owned by the Saudis, and the relationship of the small industries is positive with the suppliers, which are significant indications for potential growth.

The studies also highlight the difficulties SMEs in Saudi Arabia face, such as red tape, lack of funding, access to financing, inadequate training, and a hostile business environment. the world's e-commerce presents challenges to Saudi Arabia's SMEs, and there is a favourable correlation between e-commerce and the growth of Saudi Arabia's SMEs sector. Sadi and Henderson [9] observed in their research that franchising is a good way for SMEs to gain profit, while some problems are arising between the relations of franchisors and franchisees. They further suggested that policymakers of Saudi Arabia must consider SMEs and franchisees while formulating policies for the development of the economy.

Moreover, various regional, national, international, and enterprise factors affect the performance of the SMEs. El-Khasawneh [10] identified that the enterprises' factors are controllable, while national or external factors are the factors that can be controlled politically or through the government system or change in regulations. International factors are unpredictable and fully uncontrollable and can only be adjusted. Thakkar, Kanda, and Deshmukh [11] advocated the importance of supply chain management, which is becoming important for SMEs, providing the ability to product design and new processing technology through faster access [12].

2.4 Sectoral Employment Analysis for Saudi Arabia

To better understand Saudi industrial production, this study examines how output and pay affect labor demand at the sectoral level. The authors evaluate recent research on Saudi Arabia's employment determinants in order to set the stage for their study, concentrating on the most common factors such labor costs (i.e., wages) and economic output. Other explanatory variables mentioned include currency rates, inflation, trade, oil prices, and taxation.

The authors next assess numerous research on employment in Saudi Arabia, finding that few have examined wages as a driver of employment. The authors also point out that previous research has focused on output increase rather than output level. The authors look at research on the relationship between employment and other variables such financial market development, oil prices, government spending, education spending, FDI, inflation, and exports. However, none of these studies examined how wages affect labor demand.

The authors conclude that their study adds to the literature on industrial production in Saudi Arabia by analyzing the impact of output and wages on labor demand in 10 sectors of the economy. They use cointegration and equilibrium correction methods to analyze time-series data from 1995 to 2016, considering structural breaks in the data. The authors find that in the long run, employment is positively affected by output but negatively affected by wages in all sectors. In the short run, employment growth is influenced by wage growth in all sectors, except for the government sector. The authors also find that employment can adjust to the desired equilibrium level in all sectors, but the time horizon for the adjustment varies across sectors. A "one-size-fits-all" strategy, according to the authors, would ignore sectoral variety, hence sector-specific techniques are required.

Overall findings provide insight into the variables affecting labor demand in Saudi Arabia's industrial production sector. The authors close a gap in the literature by examining the hitherto unresearched effect of pay on labor demand. The authors provide policymakers with information on how economic production and wages influence employment by evaluating the long-run and short-run dynamics of labor demand. The report emphasizes the need of sector-specific strategies, as each sector has unique features that must be considered when developing policies to enhance Saudi industrial output.

The study also adds to our understanding of employment by confirming that economic output and wages are the most important predictors of employment, as previously proven in other economies' studies. The authors' use of cointegration and equilibrium correction methods adds to the methodological literature on analyzing the dynamics of labor demand. The study's findings have important implications for policymakers in Saudi Arabia and other economies that seek to promote industrial production and employment [13].

Table 1 compares previous industry production studies on the industrial electricity, energy demand across different regions and notes the methodological approaches used in the existing literature. The reported elasticities show significant variation, which may be due to differences in data sets, methodology, time spans, or country-specific economic structures. Sectoral employment analysis and Industrial development. However, there is no study investigating industrial electricity demand at the regional level in Saudi Arabia, and the existing studies only consider total demand with a constant price elasticity assumption and ignore the role of technological innovation in modeling.

Table 1:Previous industry production studies

Study Name	Objectives	Dataset	Methods	Results	Limitations
Modeling and Forecasting Industrial Electricity Demand for Saudi Arabia: Uncovering Regional Characteristics	Investigate Saudi Arabia's industrial electricity consumption at the regional level	- SEC via SAMA annual data from 1990 to 2019. - KAPSARC data portal.	STSM and TVCC.	Total demand forecasted to be 82.5 TWh.	Data limitations, national values are used as an approximation because the COA represents the largest weight for these industries.
Modeling Industrial Energy Demand in Saudi Arabia and Understanding Its Drivers	Decompose the change in energy consumption between an end year and a base year into drivers.	-Data obtained from the IEA (2018a). - CEIC Data (2018).	Decomposition of the estimated energy demand equation	The decomposition results showed that the activity effect was the primary driver.	Limited availability of energy-related data.
Industrial development in Saudi Arabia: disparity in growth and development	Identify well-performing and poor-performing manufacturing industries in Saudi Arabia and suggest measures to promote the growth and development of deprived or underdeveloped industries	Secondary data available on the websites of the Saudi Government	Secondary data analysis of government data on the manufacturing sector in Saudi Arabia.	Earning capacity and capital formation of different manufacturing industries in Saudi Arabia are growing at uneven and sometimes negative rates, indicating uneven development of the manufacturing sector.	Narrow scope of the study due to the limited financial variables considered

Sectoral employment analysis for Saudi Arabia	Analyze the impact of output and wages on labor demand in Saudi Arabia's ten sectors, using cointegration and equilibrium correction methods to inform sector-specific policies and promote industrial production	data provided SAMA, Annual Report, 2016	Malmquist productivity index (MPI), which is a non-parametric alternative to parametric frontier production analysis.	Saudi manufacturing sector improved in scale efficiency change and pure technical change, with efficiency change and technical change also improving. The Malmquist TFP index showed an overall improvement in productivity change	the rate of technological progress was not good for most of the examined industries, study only uses the Malmquist productivity index to measure total factor productivity.

Our research aims to overcome the limitations of previous studies by addressing all the issues. We are also aware that only one recent study by [3]. has made a projection of industrial electricity demand. However, this study only forecasted industrial electricity demand and did not forecast all industrial production index (General IPI), Electricity and gas (IIPE), mining& quarrying (IIPQ), and manufacturing (IIPM). Therefore, our study aims to make forecasts for future industrial production at the level of the four industrial sectors of Saudi Arabia, by utilizing real-data applications and numerical simulations. Furthermore, we intend to investigate several potential factors that could influence the forecasting performance of our project.

3 Data Preprocessing

The following section explains how we have conducted the preprocessing of the data where the process involves cleaning, transforming and organizing the data to ensure that it's ready for analysis. The first step was to understand the data, and the figures below depict our two data sets, which are: percentage change in the Industrial Production Index (IPI) by economic activity and the economic ratio.

	Economic Activities	Time Period	Percentage change in the Industrial Production Index
0	Mining and Quarrying	2021 / 01	-6.23
1	Manufacturing Industry	2021 / 01	-10.62
2	Industrial Production Index	2020 / 01	-6.68
3	Industrial Production Index	5/1/20	-15.52
4	Industrial Production Index	6/1/20	-22.24

Figure 1 the Industrial Production Index (IPI) by economic activity

The first figure illustrate the data set of the Industrial Production Index (IPI) by economic activity, while the second figure illustrates the economic ratio.

	Time Period	Economic Dependency Ratio
0	2017 / Q2	145.0
1	2019 / Q3	127.0
2	2016 / Q4	135.7
3	2017 / Q3	142.3
4	2018 / Q1	139.1

Figure 2 the economic ratio

The second step which was cleaning the data ensuring the data was clear from any missing value as it shows below in the figure both data sets didn't have any missing value.

```
# 2. Data cleaning
# Check for missing values in the IPI data
print(IPI_data.isnull().sum())
```

```
Economic Activities                                 0
Time Period                                         0
Percentage change in the Industrial Production Index    0
dtype: int64
```

```
# Check for missing values in the dependency data
print(dependency_data.isnull().sum())
```

```
Time Period                   0
Economic Dependency Ratio     0
dtype: int64
```

Figure 3 Cleaning data: checking for any missing values

Furthermore, we have ensured to remove any duplicates that was found in the data. the data was found to be clean therefore, there wasn't the need of conducting any further preprocessing.

4 Data Profiling

In the following section, we performed data profiling, which is the process of evaluating data and identifying the characteristics of data sets in order to get insight into the data's structure. We have evaluated the data sets separately starting first with Industrial Production Index (IPI) by economic activity.

1) Industrial Production Index (IPI) by economic activity

Starting with the static report as it shows in the figure below where the mean is 1.41, minimum is -32.11 and the maximum is 55.05.

	Percentage change in the Industrial Production Index
count	204.000000
mean	1.410833
std	13.985357
min	-32.110000
25%	-8.970000
50%	-0.740000
75%	11.492500
max	55.050000

Figure 4 IPI: static report

For the data visualization using two types to illustrate the data distribution for the IPI the first one is the histogram the other one is box plot.

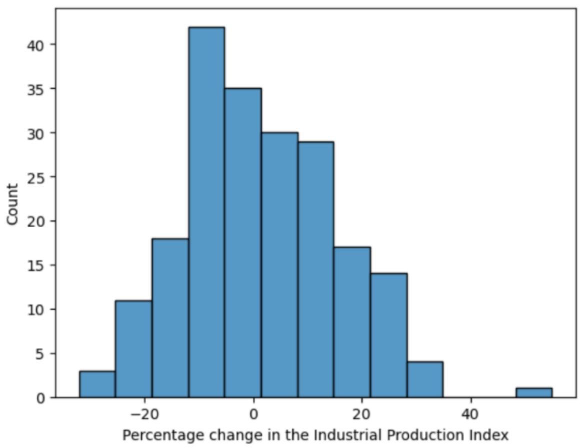

Figure 5 IPI: Histogram

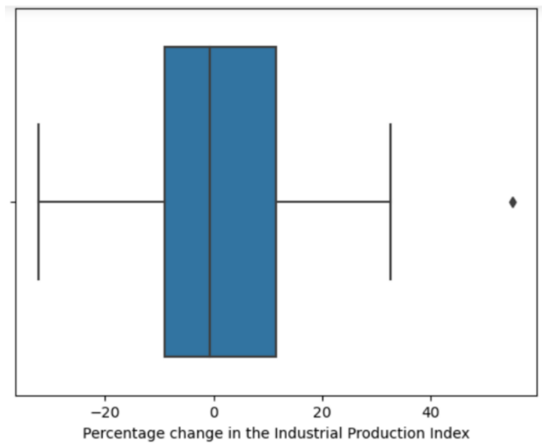

Figure 6 IPI Box plot

These two figures illustrate that there is a decrease in industrial production which indicates a decline in the production outcomes of the industrial sector.

2) Economic ratio

The second data that we have evaluated which is the economic ratio dependency, starting with the static report as it shows in the figure below where the mean is 137.05, minimum is 124.7 and the maximum is 152.6.

	Economic Dependency Ratio
count	16.000000
mean	137.050000
std	7.549834
min	124.700000
25%	133.475000
50%	136.650000
75%	141.475000
max	152.600000

Figure 7 Economic ratio: static report

Same as the IPI we have used histogram and the box plot to illustrate the data distribution.

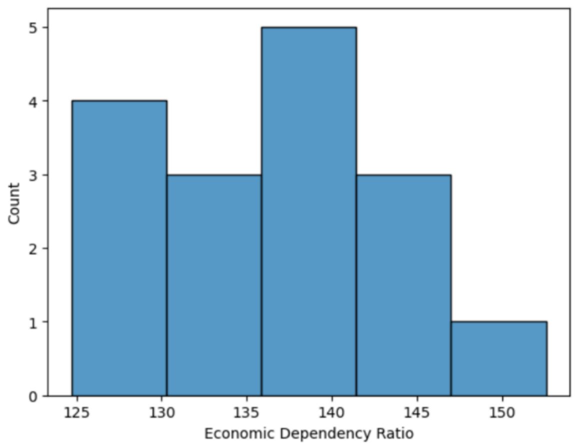

Figure 8 Economic Ratio histogram

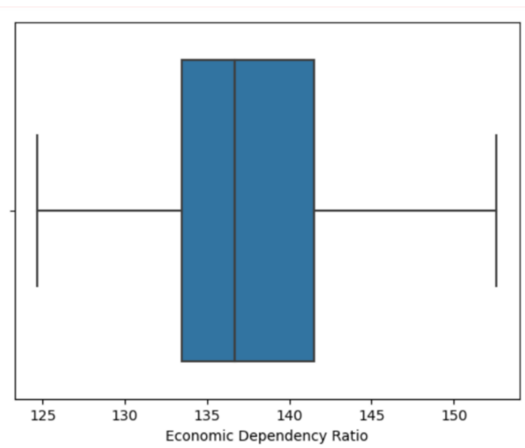

Figure 9 Economic ratio box plot

In the following figures, it signifies an increase in the economic ratio, indicating that some people are not contributing to the economy.

This study examines the relationship between the industrial indicator production and the economic ratio, where an increase in the economic ratio implies that the economy is struggling to support a growing share of unemployed people.

5 Methodology: Solution

5.1 Data Mining Model

5.1.1 Decision Tree Model

A decision tree model is a supervised learning algorithm that makes predictions using a tree-like structure. The tree is constructed by iteratively dividing the data into smaller and smaller subsets until all data points in a subset belong to the same class. The decision tree classifier is a powerful algorithm for solving classification problems. We used The Industrial Production Index (IPI) dataset to predict whether the percentage change in the Industrial Production Index (IPI) would be positive or negative using a decision tree model.

The Industrial Production Index (IPI) dataset characteristics:

Economic Activity: The category of economic activity, such as construction, transportation, or manufacturing.

Year: The year in which the observation was made.

Month: The month in which the observation was made.

The variable of interest is the class, which is either "Positive Change" or "Negative Change."

1. **80**/20 percentage split

```
Accuracy: 0.8048780487804879
Classification Report:
                 precision    recall  f1-score   support

Negative Change       0.77      0.85      0.81        20
Positive Change       0.84      0.76      0.80        21

       accuracy                           0.80        41
      macro avg       0.81      0.81      0.80        41
   weighted avg       0.81      0.80      0.80        41
```

Figure 10: Decision Tree Model 80/20 percentage split

2 90/10 percentage split

```
Accuracy: 0.8095238095238095
Classification Report:
                 precision    recall  f1-score   support

Negative Change       0.83      0.83      0.83        12
Positive Change       0.78      0.78      0.78         9

       accuracy                           0.81        21
      macro avg       0.81      0.81      0.81        21
   weighted avg       0.81      0.81      0.81        21
```

Figure 11: Decision Tree Model 90/10 percentage split

3 70/30 percentage split

```
Accuracy: 0.8225806451612904
Classification Report:
                 precision    recall  f1-score   support

Negative Change       0.84      0.81      0.83        32
Positive Change       0.81      0.83      0.82        30

       accuracy                           0.82        62
      macro avg       0.82      0.82      0.82        62
   weighted avg       0.82      0.82      0.82        62
```

Figure 12: Figure 2: Decision Tree Model 70/30 percentage split

The Decision Tree Classifier demonstrated a great accuracy value of 82.258% using 70/30 training to testing ratio. Furthermore even 90/10 great accuracy with score of 80.95%. On the other hand, the Decision Tree Classifier performed the worst when using an 80/20 ratio, with an accuracy rate of 80.487%.

5.1.2 Arima Model

For forecasting the economic dependcy ratio we have ARIMA model to predict the forcast values of the ecnomic ratio on a certain time of period. The following steps will explain how the ARIMA model was done:

1- We first started with uploaded the data using pd.read_csv().

2- We have used the 'Time Period' column is an interpretion to datetime and assigned it as the DataFrame index, with the DataFrame sorted in ascending order by the index.

3- The order (p, d, q) of an ARIMA model is (1, 1, 1), indicating that the model will have one autoregressive term, one differencing term, and one moving average term.

4- The model is fitted to the data.

5- We have then specified the number period to five quarters and uses the get_forecast() method to obtain the forecasted values and accompanying confidence ranges.

6- The projected values are extracted, and a new index is produced for them based on the original data's last index.

7- As you can see in the figure below the forcasted value are printed.

8- Using the matplotlib software, the economic dependcy ratio data and forecasted values are plotted on a graph.

5.2 Evolution Metrics

Confusion matrices are an effective technique for assessing a machine learning model's performance. They provide a more complete evaluation of the model's performance than simple accuracy ratings and can be used to detect problem areas.

The Confusion matrices show four different categorization outcomes that can be achieved by a supervised classification model.

Table 2: confession Matrix

<table>
<tr><th colspan="4">Really is</th></tr>
<tr><th rowspan="3">Classified as</th><td></td><th>Positive</th><th>Negative</th></tr>
<tr><th>Positive</th><td>TP</td><td>FN</td></tr>
<tr><th>Negative</th><td>FP</td><td>TN</td></tr>
</table>

- True positives (TP) are instances in which the model forecasted the positive class accurately.
- False positives (FP) are instances where the model forecasted the positive class wrongly.
- True negatives (TN) are cases in which the model forecasted the negative class accurately.
- False negatives (FN) are instances where the model forecasted the negative class wrongly.

To evaluate the performance of a machine learning model, a confusion matrix produces the four-evaluation metrics described below:

- **Accuracy:** is defined as the ratio of the model's correct predictions to the total number of forecasts. It is frequently given as a percentage, with a higher percentage indicating that the model makes more accurate predictions.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

- **Recall**: It is commonly referred to as sensitivity or true positive rate, and it represents the ratio of true positive forecasts to total positive cases in the data. A high recall value indicates that the model generates few false negatives.

$$Recall = \frac{TP}{(TP + FN)}$$

- **Precision:** This parameter represents the proportion of instances that were correctly predicted by the model that were positively identified. It quantifies the model's accuracy in making positive predictions.

$$Precision = \frac{TP}{(TP+FP)}$$

- **F1-Score:** Also known as the harmonic mean, this statistic assesses precision and recall in a balanced manner. It integrates both measures into a single score to provide an in-depth assessment of the model's performance.

$$F1 = \frac{(2 \times P \times R)}{(P + R)}$$

6 Experimental results: Evaluation of proposed model

6.1.1 Decision Tree Model

For the Decision Tree model, we employed the classification techniques discussed previously and implemented them with Python Jupyter Notebook. Various percentage splits, including 70%, 80%, and 90%, were used to evaluate the performance of the methods. The experiment's results, rounded to the nearest decimal place and presented in the Table below:

Table 3: Decision tree model Results

Train/ Test	70/30				80/20				90/10			
Matrix/ Change type	Accuracy	Precision	Recall	F-score	Accuracy	Precision	Recall	F-score	Accuracy	Precision	Recall	F-score
Positive change	82%	84%	81%	83%	80%	77%	85%	81%	81%	83%	83%	83%
Negative change		81%	83%	82%		84%	76%	80%		78%	78%	78%

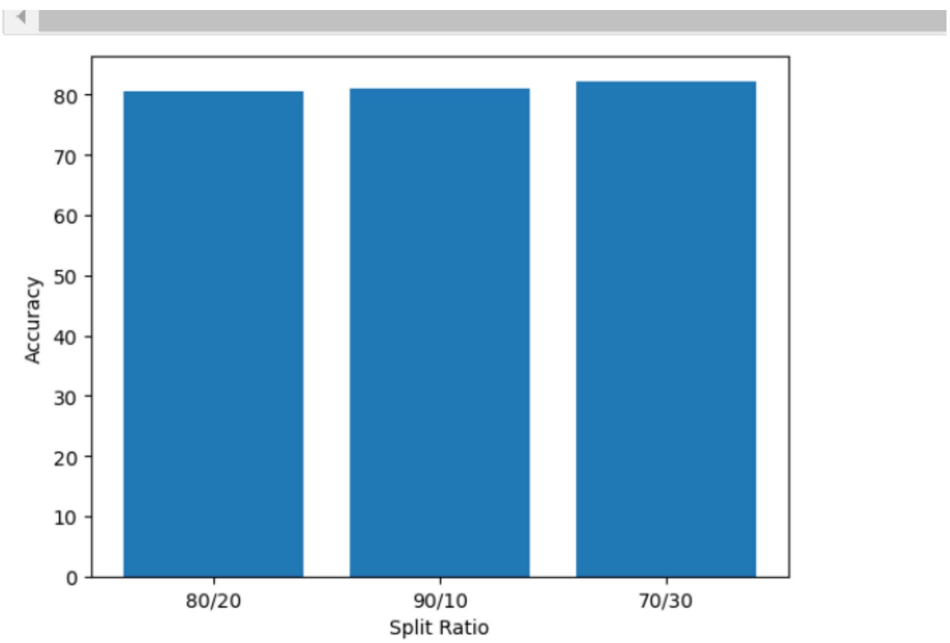

Figure 13: Results of decision tree accuracy classification

The analysis demonstrated that in Decision Tree model 70/30 training to test split achieved the highest accuracy rate of 82%. In addition, even a score of 90/10 with an accuracy of 81% is impressive. In contrast, the Decision Tree Classifier performed the worst with an 80/20 ratio, achieving an accuracy rate of 80%. However, there is no significant difference between the three experiments.

6.1.2 Arima Model

The figure below graph depicts historical data as well as anticipated numbers for the next five quarters. The time period, measured in quarters, is represented by the x-axis, while the economic reliance ratio is represented by the y-axis. The blue line indicates historical data values, while the orange line reflects anticipated values. The dark area around the orange line reflects the anticipated value' 95% confidence interval. The graph suggests the increase in the economic ratio as shown in the historical data starting from 2016 till 2020 and that this drift will continue on for the next five quarters.

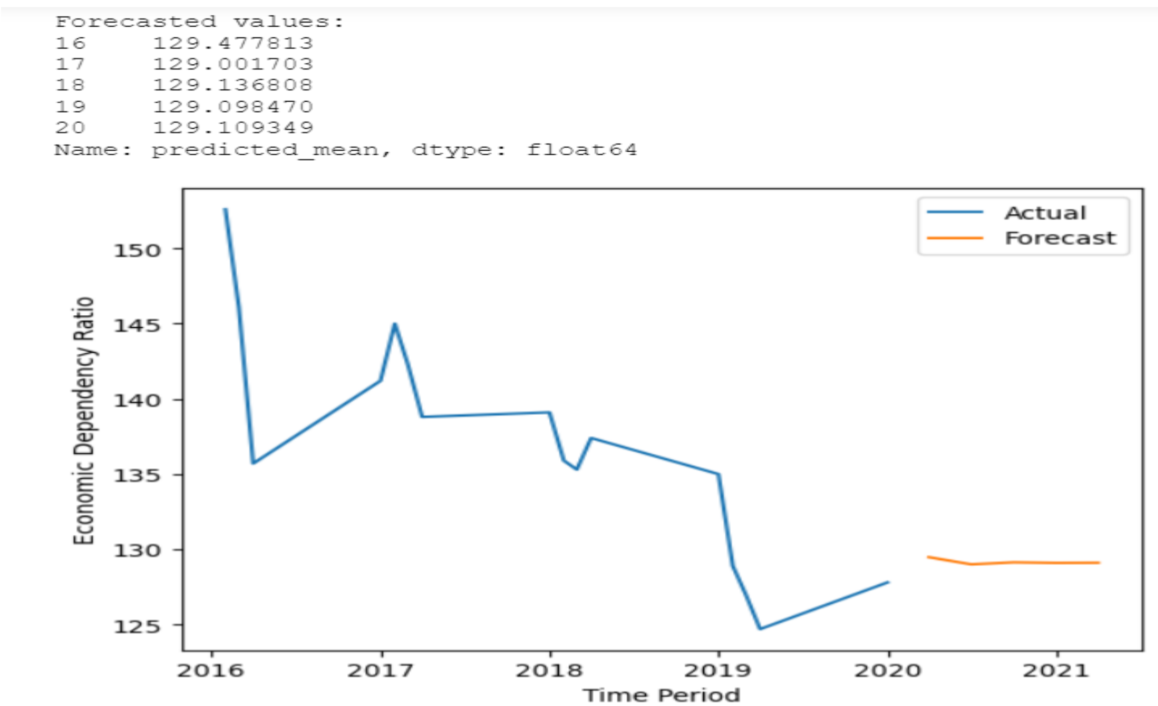

Figure 14: Economic ratio: ARIMA Forecasting model

7 Conclusion and discussion

The study's main objectives were to examine the Kingdom of Saudi Arabia's industrial production index (IPI) and create a model to anticipate its future growth. For politicians, economists, and business experts, it is essential to precisely estimate and analyze industrial output since it is so important to the nation's economy.

For accurate forecasting and analysis of industrial production in Saudi Arabia, the study revealed a number of critical challenges that must be resolved. These include identifying and simulating seasonal and cyclical patterns, taking volatility and external factors into consideration, and undertaking a sector-specific analysis that considers the specifics of each sub-sector.

The study used a mix of data mining approaches, such as a decision tree model and an auto-regressive integrated moving average (ARIMA) model, to solve these difficulties. These models were used to the Industrial Production Index dataset to anticipate future values and predict the percentage change in the IPI.

The study's conclusions showed that the industrial output index could be accurately predicted by both the decision tree model and the ARIMA model. The decision tree model distinguished between positive and negative percentage changes in the IPI, offering insightful information about the course of industrial output. On the other hand, the ARIMA model predicted the IPI's future values, enabling a better comprehension of the course and prospects of the industrial sector.

Policymakers, economists, and investors can use the study's findings and a visualization of the anticipated values for the subsequent six months to make critical decisions about resource allocation, industrial planning, and policy. Stakeholders can discover growth possibilities, reduce risks, and assist Saudi Arabia's overall economic development and diversification goals, as described in Vision 2030, by properly estimating industrial production.

It is crucial to keep in mind, however, that precise industrial production forecasting and analysis depend on the availability of up-to-date data and the evaluation of outside elements like global economic trends and geopolitical events. As a result, ongoing efforts should be undertaken to enhance data gathering and guarantee that pertinent external variables are included in forecasting models.

In conclusion, this research adds to our understanding of Saudi Arabia's industrial sector and offers useful information to economists, policymakers, and business leaders. Stakeholders can assist the long-term growth of the industrial sector and help the Kingdom achieve its overall economic development and diversification goals by solving the difficulties in forecasting and assessing industrial production.

8 References

[1] GASTAT Statistical Database, https://database.stats.gov.sa/ (accessed May 17, 2023).

[2] "Saudi Arabia Industrial productionapril 2023 data - 1963-2022 historical," Saudi Arabia Industrial Production - April 2023 Data - 1963-2022 Historical, https://tradingeconomics.com/saudi-arabia/industrial-production (accessed May 18, 2023).

[3] J. I. Mikayilov *et al.*, "Modeling and Forecasting Industrial Electricity Demand for Saudi Arabia: Uncovering Regional Characteristics," *KAPSARC*, Dec. 2021. [Online]. Available: https://www.kapsarc.org/research/publications/modeling-and-forecasting-industrial-electricity-demand-for-saudi-arabia-uncovering-regional-characteristics/

[4] S. Alarenan, A. Gasim, and L. C. Hunt, "Modeling Industrial Energy Demand in Saudi Arabia and Understanding Its Drivers," *KAPSARC,* Dec. 2019. [Online]. Available: https://www.kapsarc.org/research/publications/modeling-industrial-energy-demand-in-saudi-arabia-and-understanding-its-drivers/. [Accessed: May 16, 2023]

[5]" Ahmed Al Bakr" www.sama.gov.sa. https://www.sama.gov.sa/en-US/EconomicResearch/WorkingPapers/Challenges%20to%20Production%20Base%20Diversification%20in%20Saudi%20Arabia.pdf (accessed May 19, 2023).

[6] R. N. Kayed and M. Kabir Hassan, "Saudi Arabia's economic development: entrepreneurship as a strategy," *International Journal of Islamic and Middle Eastern Finance and Management*, vol. 4, no. 1, pp. 52–73, Apr. 2011, doi: https://doi.org/10.1108/17538391111122203.

[7] Y. Mehta and A. J. Rajan, "Manufacturing Sectors in India: Outlook and Challenges," *Procedia Engineering*, vol. 174, pp. 90–104, 2017, doi: https://doi.org/10.1016/j.proeng.2017.01.173.

[8] A. S. Maghrabi, Y. A. Jefery, and S. A. Bin Sabbaan, "The market competition of Saudi small-sized manufacturing firms: an exploratory study," *International Journal of Commerce and Management*, vol. 19, no. 4, pp. 309–320, Nov. 2009, doi: https://doi.org/10.1108/10569210911008485.

[9]M. Asad Sadi and J. C. Henderson, "Franchising and small medium-sized enterprises (SMEs) in industrializing economies," *Journal of Management Development*, vol. 30, no. 4, pp. 402–412, Apr. 2011, doi: https://doi.org/10.1108/02621711111126855.