



Article

# Contextual Intelligence: An AI Approach to Manufacturing Skills' Forecasting

Xolani Maphisa, Mpho Nkadimeng and Arnesh Telukdarie \*

Johannesburg Business School, University of Johannesburg, Johannesburg 2092, South Africa;  
gnkadimeng@hotmail.com (M.N.)

\* Correspondence: arnesht@uj.ac.za

**Abstract:** The manufacturing industry is skill-intensive and plays a pivotal role in South Africa's economy, reflecting the nation's progress and development. The advent of technology has initiated a transformative era within the manufacturing sector. Workforce skills are at the heart of ensuring the sustained growth of the industry. This study delves into the skill-related aspects of the occupational landscape of the South African manufacturing sector, with a particular focus on two important manufacturing sectors: the food and beverage manufacturing (FoodBev) sector and the chemical manufacturing (CHIETA) sector. Leveraging the forecasting prowess of Autoregressive Integrated Moving Average (ARIMA), this paper outlines a sectorial occupational forecasting modeling exercise to reveal which job roles are poised for expansion and which are expected to decline. The approach predicted future skills' demand 80% accuracy for 473 out of 713 (66%) occupations for FoodBev and 474 out of 522 (91%) for CHIETA. These insights are invaluable for industry stakeholders and educational institutions, providing guidance to support the sector's growth in an era marked by technological advancement.

**Keywords:** ARIMA forecasting; labor market; data-driven decision-making



**Citation:** Maphisa, X.; Nkadimeng, M.; Telukdarie, A. Contextual Intelligence: An AI Approach to Manufacturing Skills' Forecasting. *Big Data Cogn. Comput.* **2024**, *8*, 101. <https://doi.org/10.3390/bdcc8090101>

Academic Editor: Domenico Ursino

Received: 8 July 2024

Revised: 2 August 2024

Accepted: 13 August 2024

Published: 2 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The manufacturing industry stands as an undeniable backbone in South Africa's economic framework, not merely serving as a gauge of the nation's progress but also as a driving force behind its economic prosperity. Of these industries, the food and beverage manufacturing (FoodBev) sector plays an especially pivotal role, ensuring that South African households enjoy a consistent supply of a diverse range of sustenance. Its impact extends well beyond the nation's borders, substantially contributing to exports and fortifying foreign profits. The chemical manufacturing (CHIETA) sector provides the essential foundation for various industries, including agriculture, mining, and pharmaceuticals, enabling the production of a wide array of crucial goods, from fertilizers to life-saving medications. These vital sectors have not been immune to the transformative wave brought by the recent advent of the 4th industrial revolution. Within this profound shift, the skills of the workforce have risen to prominence as a critical determinant for the industry's sustained growth. Understanding the current scarcity of specific occupations and forecasting those set to flourish in the future holds great significance in the manufacturing sector. The overall purpose of forecasting is to conclude what occupations will be required by the labor market on the selected horizon in a given sector [1]. The availability of employment forecasts serves as an invaluable early-warning system for potential skill and job shortages, offering the manufacturing sector and its associated training providers the opportunity to adjust the supply of skills required to fulfill certain occupations. This proactive approach can help mitigate the detrimental effects of skill shortages and ensure a more robust and resilient manufacturing industry in South Africa.

Anticipating the changes in occupation demand has proven to be a formidable challenge, given its dependence on an array of factors, including technological advancements,

shifts in industrial structure, and fluctuations in economic activity [2]. Numerous approaches have emerged over time to address the complex task of forecasting employment rates. Early research predominantly favored quantitative methods, primarily because the provision of quantitative results was deemed essential to meet the needs of potential users of the forecasts [3]. The availability of data is a critical factor in the accuracy of quantitative forecasting techniques [4]. Naturally, the complexity of the forecasting problem influences the technique selected. In instances where the available data are insufficient to support these quantitative methodologies, alternative approaches have come to the fore. Techniques such as the Delphi method [5], harnessing the collective expertise of industry professionals [3], along with qualitative surveys and interviews [6], offer valuable qualitative insights. Additionally, literature review [7] and content analysis [8] have been adopted to extract information from existing sources. While these non-quantitative approaches provide useful supplementary insights, they are generally seen as complements to, rather than substitutes for, fully fledged quantitative-based projections.

Historically, South Africa has seen limited attempts in occupational forecasting, often reliant on qualitative labor market assessments due to constraints in data quality and availability [3,9,10]. In the early 2000s, the Human Sciences Research Council (HSRC) conducted occupational forecasts [9,10]. As noted in [3], these studies were heavily limited by data availability; hence, they were restricted to a qualitative assessment. A transformative shift emerged with the introduction of the Sector Skills Plan (SSP) framework by the Sector Education and Training Authorities (SETAs) in 2015. The SSP framework mandated SETAs to produce documents outlining skills, employment profiles, and training interventions within their corresponding economic sectors, drawing data primarily from the Work Skills Plan (WSP). Under this framework, businesses affiliated with SETAs must submit WSPs, creating a rich data repository encompassing occupational profiles, skill demands, shortages, and training initiatives. SETAs manage the WSPs from the businesses within their jurisdiction, offering a comprehensive snapshot of employment within specific economic sectors, such as FoodBev and CHIETA manufacturing sectors. Leveraging this extensive and invaluable dataset opens the door for the development of a quantitative forecasting model—an unexplored territory within South Africa's forecasting landscape to date.

Globally, sectoral bodies have been established to promote skills' development. Countries such as Canada, the USA, and New Zealand have sectoral bodies that conduct occupational forecasting using Manpower Requirements Approach (MRA)-based forecasting techniques. These projections form the basis for the countries' decision-making process for strategic workforce development [1,11]. While the SETAs perform significant work in skills' development, they do not have a model in place for occupational forecasting. This study aims to fill this gap by developing an occupational forecasting model. Using the WSP data provided by FoodBev and CHIETA, this study employs the ARIMA forecasting model to explore the occupational landscape within the two SETAs.

The study offers several key contributions. Firstly, it introduces a robust forecasting model tailored to the South African context, addressing a critical gap in the current skill development framework of the SETAs. Second, the study offers a methodological contribution by demonstrating the application of ARIMA in occupational forecasting, which can be adapted and applied to other SETAs. By achieving these aims, the study seeks to provide actionable insights that can guide strategic workforce planning in South Africa, ensuring that the workforce is well equipped to thrive under the current industrial revolution.

## 2. Literature Review

### 2.1. Manufacturing and Technology with Dependency on Skills

Industry 4.0 is driving the digitization of the manufacturing sector, leading the development of smart products, machines, processes, and factories [12]. This transformation involves the application of cyber physical systems (CPS), supported by technologies such as internet of things, big data, cloud computing, and additive manufacturing technologies, to name a few. The integration of these novel technologies is reshaping the manufacturing

landscape, altering processes, skill requirements, and occupational profiles needed to perform tasks [13]. Ref. [14] highlighted that the previous industrial revolutions significantly altered occupational profiles, transforming employee roles and necessary skills. The disruptive impact of Industry 4.0 resulted in work processes undergoing significant changes, necessitating an adjustment to how work is performed [15]. Ref. [16] further highlighted that technology will significantly transform employees' work profiles. The future factory will feature a significant presence of collaborative robotics capable of interacting with humans in the workplace. This suggests that, while the extent of automation will differ across various occupations, its impact will be widespread [17]. In such situations, humans will need to acquire logical skills that complement advanced robots [18]. The literature consistently indicates that there will be an increase in automation and advanced robotics taking over routine and repetitive tasks. The work in [19] examined which occupations are susceptible to automation. A total of 702 occupations were examined and 47% of jobs were classified as highly susceptible to automation, particularly those that include repetitive tasks. Ref. [20] highlighted that the decreasing employment rate in the manufacturing industry is due to the reduction in routine jobs. Manufacturing occupations typically involve tasks that follow a clearly defined repetitive process, which can now be encoded into a software program and, therefore, executed by a computer [21].

In contrast, Ref. [22] suggests that the rising automation should not be viewed as a threat but rather as an opportunity: workers will be liberated from repetitive tasks to focus on areas where they can add significant value. This perspective suggests that new technologies could have a positive impact on employment by creating demand for a wide range of skills, including those needed for managing Industry 4.0 technologies. Building on this idea, Ref. [23] proposed three main outcomes of technological innovation: skills that compete with automation will be reduced, skills that complement automation will increase, and finally, skills where machines fall short will increase.

Considering these transformative shifts and the anticipated change in skill demands, the necessity of forecasting occupations becomes evident. Forecasts serve as a strategic tool, offering insights into the evolving job landscape, informing workforce planning, and preparing individuals and industries for the skills essential in a technologically evolving world.

## 2.2. Occupational Forecasting—International Perspective

A wide range of techniques for skills and occupational forecasting have been explored worldwide. However, current efforts in this field remain heavily constrained by data limitations. The feasibility of different forecasting methods is largely dependent on the data infrastructure available in each country. Nations such as the USA, Canada, and European countries have been in the arena of occupational forecasts for several decades. Their advanced analyses are supported by significant investments in data gathering and modeling capabilities. Large databases have been established over the years, and this significantly aids in building robust and more informed forecasting models.

In the USA, the Bureau of Labor Statistics (BLS) has been a key player in occupational projections since the 1940s, employing an elaborate methodology based on industry-specific occupational requirements [24]. The BLS derives its projections from the basic model, where occupational requirements are estimated for each industry based on projected output growth, growth in labor productivity, and the occupational composition of each industry. These requirements are then aggregated to produce occupational requirements for the economy as a whole. This methodology has been continually refined by researchers [25–27] over the years. Alongside the BLS, the O\*NET database, characterized by standardized occupation descriptors, is a valuable resource updated regularly from input across various occupations. Utilizing data from both BLS and O\*NET, Ref. [28] employed machine learning models to predict growing and declining occupations with increased precision, demonstrating the potential of these rich databases in enhancing forecasting accuracy.

In the European context, the European Centre for the Development of Vocational Training (CEDEFOP) plays a significant role in developing occupational forecasts for individual countries and groups of countries within the European Union [29,30]. The CEDEFOP skills forecast provides quantitative estimates for future employment trends across different economic sectors and occupational groups. The adoption of the International Standard Classification of Occupations (ISCO) in continental Europe has facilitated inter-country comparisons and alignment of forecasting outcomes across various European nations. This standardized classification framework has allowed researchers to transcend national boundaries, enabling comprehensive assessments and cross-country analyses of occupational forecasts, fostering a unified approach to understanding future occupational trends within the region.

In the last 30 years, the Canadian Occupational Projection System (COPS) has been employed in Canada to produce a 10-year labor market forecast every 2 years [11]. The estimation of occupational supply by COPS involves a synthesis of projections for immigrants, graduates, dropouts, and re-entrants, coupled with forecasts for labor force participation rates. Through the combination and scrutiny of these demand and supply projections by occupations, the COPS model discerns whether the future labor market is in balance or if certain occupations will encounter shortages or surpluses.

### 2.3. Occupational Forecasting—Local Perspective

The landscape of advanced occupational forecasting in South Africa is a work in progress, yet it has seen notable attempts to predict the future labor market. In 1999, the Human Sciences Research Council (HSRC) conducted extensive research of South African labor market trends, analyzing formal employment within eight economic sectors over a five-year period, excluding the agricultural sector [9]. This comprehensive study predicted future demand for various employment roles. The research initially involved a survey of 273 companies to collect data on current employment, projected supply, skill demand, and anticipated shifts in future skill requirements, culminating in the development of a comprehensive demand forecasting model for 1998 to 2003.

Subsequently, in 2001, a commissioned study by the European Union, the Department of Labor, and the Department of Trade and Industry aimed to investigate critical skill shortages and expedite skills' development [3]. This multifaceted study utilized a blend of qualitative, quantitative, and meta-analytical techniques. The results revealed a significant increase in High-Level Human Resource (HLHR) occupations in the South African labor market, particularly from 1965 to 1994. This growth was especially notable in occupations such as engineers, accountants, managers, and IT-related roles.

In 2003, Ref. [10] extended the previous research in [9], providing updated labor market projections for certain occupations from 2001 to 2006. Their approach involved using a labor demand model to estimate new positions resulting from sectoral growth and a distinct "replacement demand" model to determine demand due to retirements, emigration, and inter-occupational mobility. Interestingly, even in occupations projected to experience substantial declines in employment levels, the need to train new individuals was emphasized to maintain the existing stock of skills at required levels.

Table 1 provides a summary of the international occupation models discussed in the previous section along with local forecasting initiatives. A critical and apparent distinction is that South Africa has seen limited forecasting initiatives. The HSRC has contributed mostly to this; however, the availability of quality employment data has seen their efforts not progress. One can notice that in the global landscape, projections are continuous, and this is due to the data infrastructures that have been put in place, allowing the development of robust forecasting models.

Previous studies attempting to forecast changing occupational demands in South Africa have consistently highlighted concerns regarding data availability and quality [3]. The restricted access to reliable data has posed a significant challenge in formulating an effective occupational model for South Africa. However, a breakthrough arrived with the

introduction of the Skills Sector Planning framework by Sector Education and Training Authorities (SETAs) in 2015, marking a crucial step forward in addressing the limitations of data availability and quality that have hampered forecasting initiatives. The primary objective of all 21 SETAs in South Africa is to facilitate skills' development by establishing a spectrum of learning programs, such as learnerships, skills programs, internships, and other strategic learning initiatives. Each SETA is entrusted with developing skills in the specific economic sector it serves. Leveraging the Skills Sector Planning (SSP) framework, SETAs are mandated to produce annual reports detailing employment profiles, skills deficits, and strategic training interventions. These reports draw from the Work Skills Plan (WSP) submitted by businesses to the respective SETAs they are associated with. The employment data encapsulated within the WSP reflect a substantial portion of the overall employment within the economic sector, rendering it a viable source for constructing a rudimentary occupational model. In this context, this research aims to explore predictive analytics techniques utilizing data from two SETAs, FoodBev and CHIETA. The objective is to develop a forecasting model, capitalizing on the data reservoirs provided by the two SETAs.

**Table 1.** Summary of occupational forecasting models.

Country	Model/Study	Capabilities/Description
USA	Bureau of Labor Statistics (BLS)	Long-term occupational projections and comprehensive economic sector analysis.
European Union	CEDEFOP	Skills forecast with quantitative estimates and cross-country analysis of occupational trends.
Canada	Canadian Occupational Projection System (COPS)	Ten-year labor market forecasts every two years. Projects labor supply and demand in order to balance potential occupational shortages or surpluses.
South Africa	Human Sciences Research Council (HSRC)—1999	Analyzed formal employment trends in eight sectors over five years and developed a demand forecasting model for 1998–2003.
South Africa	EU, Department of Labor (South Africa), and Department of Trade and Industry—2001	Investigated critical skill shortages and skills' development using qualitative, quantitative, and meta-analytical techniques.
South Africa	Updated HSRC study from 1999 to 2003	Provided updated labor market projections using labor demand and replacement models.

#### 2.4. Forecasting Techniques

Forecasting future employment trends in the labor market is a critical task, facilitated by the application of predictive analytics. This section explores commonly used algorithms for time series predictions using historical data. When deciding on a method for time series forecasting, careful consideration of the characteristics of the dataset and the forecasting horizon becomes imperative. Notable algorithms include autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA), long short-term memory (LSTM), and random forest. Because each of these methods has specific advantages and disadvantages, the applicability of a given option depends on the properties of the data and the inherent nature of the prediction problem. Several time series forecasting studies have been carried out, in which researchers compare different approaches to determine which is best appropriate in a given situation. The study in [31] examined and contrasted three models' modeling and predicting capabilities in relation



to seasonal artificial neural network (SANN), SARIMA, and ARIMA models. In another study [32], the observed error between the ARIMA and the more complex SARIMA methods suggests that the performance of the two methods is nearly comparable. Additionally, while recognizing the trade-off between algorithmic accuracy, complexity, and computing speed, the authors of [33] demonstrated random forest's greater performance over ARIMA. In another study [34], random forests were identified as the top choice, whereas ARIMA also performed well. Ref. [35] noted that optimal model selection is influenced by forecast horizon, as different horizons are associated with varying data distributions. While researchers suggest different models, the prediction context and characteristic of the data stand as vital considerations. It is essential to understand these factors before selecting the most suitable forecasting model.

Among the numerous methodologies tested in various fields, the frequently used ARIMA model remains a favored option, especially for forecasting unemployment and employment rates [36–38]. Studies have illustrated its high precision, as evidenced by studies such as [39], which investigated labor market wage forecasting using advanced ARIMA functions. This indicates that, with proper settings, the ARIMA model can produce favorable outcomes in many circumstances. In [40], aimed at predicting the number of available occupations in the Russian arctic zone, exponential smoothing and neural networks were used. The study concluded that the ARIMA model demonstrated the greatest precision compared to the three other models tested. In other domains of time series forecasting, machine learning is often employed in complex areas where several challenging factors are present, as seen in energy demand forecasting studies [41,42]. There is a prevailing trend toward using deep learning models as forecasting problems become more challenging, despite the challenges of effectiveness and reproducibility that come with such models [43]. Recurrent neural networks with long-term dependency management capabilities, such as LSTM, show promise in predicting complex patterns. A recent study [44] highlighted the potential of AutoML in predictive analytics, demonstrating its efficacy in comparison to conventional ensemble learning methods and k-nearest oracle-AutoML models for predicting student dropouts in Sub-Saharan African countries. This underscores the trend toward using automated and hybrid approaches in forecasting, which can enhance predictive accuracy. However, autoregressive models are still the most popular when it comes to labor market forecasting [45]. In summary, it is important to consider different approaches to forecasting, including hybrid methods, machine learning methods, and traditional methods, such as ARIMA, to gain a better understanding of labor market forecasting and enhance the predictive accuracy of the forecasting problem.

### 3. Theoretical Framework

In the mid-twentieth century, a pivotal shift in decision systems research marked the beginning of a comprehensive exploration into decision-making processes, encompassing both human- and machine-driven choices, endowed with formidable predictive capabilities [46]. This era of inquiry laid the groundwork for understanding decision systems in the broader context of people, processes, systems, and data. Recent years have seen a remarkable convergence of advances in analytics, big data, machine learning, and data science to help navigate the intricacies of decision-making [47]. In this landscape, the prominence of data-driven decision-making (DDDM) has soared. Data-driven decision-making describes the methodical gathering, analyzing, examining, and interpreting of data to make well-informed judgments, which is performed by applying analytics or machine learning methodologies and techniques [48]. This approach stands as a beacon for delivering more informed and high-quality decisions by harmonizing the intuition and experience of human decision-makers with the analytical power of data. Scholars [46,47] have championed DDDM as a transformative solution, ushering in an era where rational choices are guided by a synergy of human expertise and data-driven analyses, promising superior outcomes. The implementation of data-driven frameworks for demand forecasting, as highlighted in [49], showcases the precision and adaptability of DDDM methodologies in practical

scenarios. The work in [50], where agent-based modeling was utilized for forecasting emerging infectious diseases, further exemplified how data-driven simulations can enhance public health strategies and decisions on emerging diseases. Additionally, the review of data-driven techniques in [51] underscores the versatility of data-driven techniques across various sectors.

The transformative impact of data-driven decision-making extends to the enhancement of decision quality, a concept elaborated in [52]. Through a deeper understanding of data, analytics, variable relationships, and resulting information, decision-makers are empowered to make more informed and higher-quality decisions. Analytics focus on atomic decisions, such as prioritization, classification, association, and filtering, producing outputs that serve as invaluable input for decision-makers. The newfound information and relationships, when acted upon, contribute to the enhancement of rational choices that align with overarching goals and yield positive outcomes [46].

In terms of forecasting occupations in the manufacturing sector, the data-driven decision-making (DDDM) theory appears as an optimal and relevant paradigm. This framework, by emphasizing informed decision-making grounded in data and analytics, contributes to the strategic alignment of skill development initiatives, thus optimizing the outcomes of training, education, and workforce planning in the manufacturing sector.

#### 4. Materials and Methods

In selecting an appropriate theoretical lens for this study, the data-driven decision-making (DDDM) theory emerged as a fitting choice. Esteemed by various researchers in educational systems, DDDM has proven instrumental in enhancing educational strategies for the future [46,52]. Positioned as a foundational framework, DDDM will aid SETAs and associated stakeholders in making well-informed decisions concerning talent acquisition, skill development, and resource allocation. This contribution, in turn, shapes optimized workforce planning strategies. For instance, the theory plays a pivotal role in preventing resource inefficiencies by avoiding an oversupply of skilled individuals, which might result from overestimating the demand for certain roles. Conversely, underestimating demand could lead to skill shortages, impacting productivity and innovation within the sector. The significance of DDDM becomes apparent in its profound influence on strategic planning [47]. The application of analytical models and tools, methodologically described in this section, enables accurate demand predictions while minimizing errors.

According to [47], DDDM is based on five main elements, as shown in Figure 1. Data and analytics belong to the modern theory of decision-making, while the last three elements belong to the classical theory of decision-making.

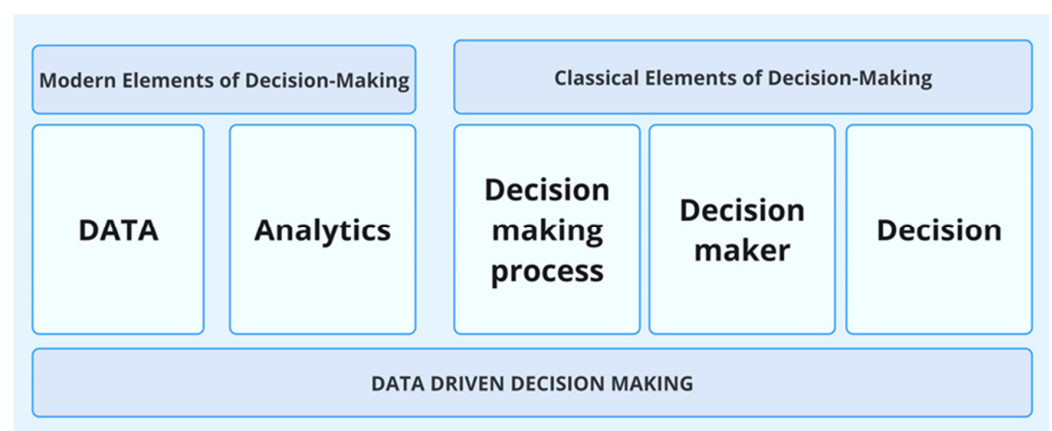


Figure 1. DDDM.

#### 4.1. Data

The foundational element involved the collection of occupational data pertinent to the FoodBev and CHIETA, spanning from 2016 to 2023. These data were sourced from the Work Skills Plan (WSP) and Annual Training Reports (ATRs). Both sectors maintain a robust data collection system, obliging firms to submit individual employee records as part of their mandatory grant applications (WSP and ATR). Impressively, the return rates for this data collection are exceptionally high, with the inclusion of employees from WSP submissions representing approximately 85% of the workforce in each sector.

The occupational data, spanning from 2016 to 2022 for FoodBev and from 2016 to 2023 for CHIETA, were obtained from the respective WSP data submissions in Excel 2016 workbooks. Before forecasting, thorough data-processing was essential. Each year's WSP data were assessed, reporting the number of employees with unique identifiers. The employee unique identifiers are a strict requirement for WSP submission; hence, no null values were identified. The unique identifiers were used to eliminate duplicates, which often occurred when large companies and their subsidiaries submitted overlapping employee information. After identifying and removing duplicates, missing values in categories such as OFO (Organizing Framework for Occupations) codes, which are used for aggregating employment numbers in each occupation, were assessed. The percentage of missing data was assessed, which was found to be low and completely random. To ensure the missing data did not affect the forecast, the Multiple Imputation by Chained Equations (MICE) technique was employed. After assessing the missing values, the quality and consistency of the data were assessed using the linear mixed-effects model. The cleaned data were then exported to a SQL database to handle future scalability and to leverage the powerful querying capabilities for complex data manipulations and aggregation. Occupations were aggregated using OFO codes. Employment numbers with significant inconsistencies, such as sudden unexplained changes from very low to very high, were identified and removed to prevent inaccurate predictions. The final dataset included 713 for FoodBev and 522 for CHIETA.

#### 4.2. Analytics

The study employed a rigorous analytical approach grounded in ARIMA models and the Box–Jenkins methodology. The selection of ARIMA models for occupational forecasting was grounded in their adeptness at handling time series data, a characteristic prevalent in workforce trends [36]. ARIMA's capacity to capture sequential dependencies and explicitly model seasonality aligns well with the nuanced nature of occupational data, where past job counts and seasonal variations significantly influence future trends [36].

The Box–Jenkins methodology, rooted in time series analysis, has gained prominence for its effectiveness in modeling and forecasting economic variables [36]. In the realm of labor economics, where understanding and predicting workforce dynamics is crucial, the application of this methodology holds significant potential. The Box–Jenkins methodology, pioneered by George Box and Gwilym Jenkins, stands as a pivotal approach in time series analysis, and within this framework, the autoregressive integrated moving average (ARIMA) model with parameters  $(p, d, q)$  has emerged as a versatile and widely applied tool [53].

The ARIMA model encompasses three key components: (1) The autoregressive (AR) component, denoted as  $p$ , which captures the linear relationship between the current observation and its past values, (2) the integrated (I) component, denoted as  $d$ , signifying the number of different operations required for achieving stationarity, and (3) the moving average (MA) component, denoted as  $q$ , which models the dependency between the current observation and residual errors from a moving average model.

Achieving stationarity is crucial in time series analysis, as a stationary time series is one whose properties do not depend on the time at which the series was observed. Time series with trends or seasonality are not stationary, and this can be addressed through dif-



ferent operations. The methodology involves a systematic process of model identification, estimation, and diagnostic checking.

The first step of the ARIMA process involved the identification of the model order ( $p$ ,  $d$ ,  $q$ ) through the analysis of autocorrelation and partial autocorrelation functions, which extended to the application of statistical tests, such as the Augmented Dickey–Fuller (ADF) test to assess stationarity and the Ljung–Box test to detect residual autocorrelation.

The autoregressive model of order  $p$  can be expressed by Equation (1), as follows:

$$Y_t = a + B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_p Y_{t-p} + \varepsilon_t, \quad (1)$$

where:

$Y_t$  = the time series,

$a, B_i$  = coefficients,

$\varepsilon_t$  = white noise.

Autoregressive models showcase their adaptability by effectively handling a wide range of time series patterns. The versatility of these models is highlighted, as different parameter values lead to the emergence of distinct and discernible patterns in the data.

On the other hand, a moving average model of order  $q$  is represented by Equation (2), as follows:

$$Y_t = a + \varepsilon_t + c_1 \varepsilon_{t-1} + c_2 \varepsilon_{t-2} + \dots + c_q \varepsilon_{t-q}, \quad (2)$$

where ( $Y_t$ ) is the time series a constant ( $a$ ) plus the moving average of current and previous white noise error ( $\varepsilon_t$ ).

Moving average models enhance their forecasting capabilities by incorporating past forecast errors in a regression-like model. This utilization of historical forecast errors contributes to the model's effectiveness in predicting future values [37]. Once the model order has been identified, involving the determination of values for  $p$ ,  $d$ , and  $q$ , the next step is to estimate the parameters  $a$ ,  $B_i$ , and  $c_i$  from Equations (1) and (2). The estimation of the ARIMA model was performed using maximum likelihood estimation (MLE), a technique aiming to find parameter values that maximize the likelihood of obtaining the observed data. MLE is akin to the least squares estimates used in regression models, minimizing the sum of squared errors, as illustrated by Equation (3):

$$\sum_{t=1}^T \varepsilon_t^2 \quad (3)$$

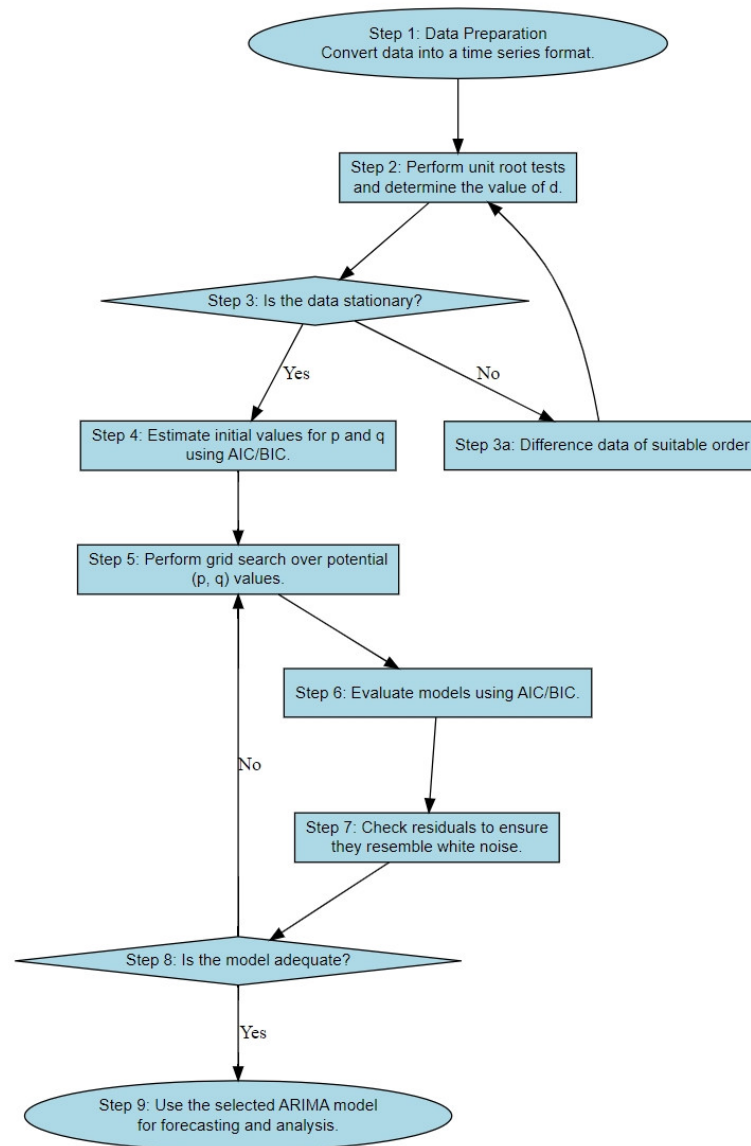
Notably, ARIMA models are more complicated to estimate compared to regression models, and different software tools may yield slightly different results due to varying estimation methods and optimization algorithms. During the estimation process, the reported log likelihood of the data represents the logarithm of the probability of the observed data arising from the estimated model. In addition to MLE, Akaike's Information Criterion (AIC) plays a pivotal role in determining the order of an ARIMA model. Analogous to its utility in selecting predictors for regression, AIC is calculated via Equation (4) as:

$$AIC = -2\log(L) + 2(p + q + k + 1) \quad (4)$$

where  $L$  denotes the likelihood of the data, and  $k = 1$  if  $c \neq 0$  and  $k = 0$  if  $c = 0$ . The last term in parentheses corresponds to the total number of parameters in the model, including  $\sigma^2$ , the variance of the residuals. For ARIMA models, a corrected version of AIC is used, denoted as  $AICc$ . Minimizing AIC,  $AICc$ , or BIC leads to obtaining well-fitted models. Subsequently, the diagnostic checking phase ensures the model's adequacy by examining residuals for autocorrelation and normality. Finally, the forecasting step utilizes the estimated model to predict future values.

#### 4.2.1. Model Validation

As a preliminary assessment before the final forecasting, the accuracy of the model was tested. The available data served as a reference for gauging the model's accuracy. This iterative refinement process continued until the optimal  $(p, d, q)$  ARIMA parameters were identified, ensuring a high level of accuracy before proceeding to forecast. Figure 2 illustrates the step-by-step process followed to select the most optimal  $(p, d, q)$  parameters.



**Figure 2.** Process diagram for  $(p, d, q)$  parameter selection.

After the initial data pre-processing described earlier, the data were converted to a time series format suitable for ARIMA modeling. This step is crucial, as ARIMA models require data to be in a sequential, time-dependent format. Following this, a stationarity check was performed using the Augmented Dickey–Fuller unit root test, to determine if differencing was needed. This led to the identification of the value of  $d$  to make the series stationary. Initial values for the autoregressive ( $p$ ) and moving average ( $q$ ) were then estimated using information criteria, specifically the Akaike Information Criterion ( $AIC$ ) and the Bayesian Information Criterion ( $BIC$ ). Several initial models were fitted, each representing unique configurations. The model with the smallest  $AIC$  value in this exploration was designated as the current model, guiding subsequent adjustments.

Subsequently, a grid search was performed over a range of potential values for  $p$  and  $q$ . This exhaustive search helped identify the combinations that yielded the best model fit. Model evaluation followed, where the fitted models were evaluated using AIC/BIC to compare their relative quality, aiming to select the model with the lowest AIC/BIC value. From the evaluated models, the one with the lowest AIC/BIC was chosen in the model fitting and selection step, ensuring that the selected model was the best fit for the data. Model diagnostics were then performed by checking the residuals of the selected model to ensure they resembled white noise, indicating no patterns, autocorrelations, or trends, thus validating the model's adequacy. Finally, the selected ARIMA model was used for forecasting and further analysis, providing the most accurate predictions based on the obtained data.

#### 4.2.2. Accuracy

The key factor for evaluating a forecasting model is accuracy, which is often the primary challenge in time series forecasting [54]. This is because it measures the level of agreement between actual and predicted values, showing the variance between them. There are several accuracy measures available for time series forecasts, such as absolute percentage error (APE), mean absolute error percentage (MAPE), and root mean squared percentage error (RMSPE). In this study, MAPE was the chosen accuracy measure. MAPE is preferred to RMSPE because it is scale-dependent and easier to interpret [55]. The MAPE was calculated by taking the sum of the absolute errors for each time period, divided by the actual value of that period, and then dividing by the number of periods, which resulted in a mean value that was converted into a percentage:

$$MAPE = \frac{\sum_{t=1}^N \left| \frac{E_t}{Y_t} \right|}{N} \times 100 \quad (5)$$

A MAPE of less than 10% is considered highly accurate forecasting, 10–20% is considered good, 20–50% is considered reasonable, and above 50% is considered inaccurate forecasting [56]. Ex-post forecasting was performed using the available data minus one year for comparison. To enhance the decision-making process for necessary interventions based on projected results, final projections were only made for occupations with a MAPE of less than 20%, which is considered good forecasting.

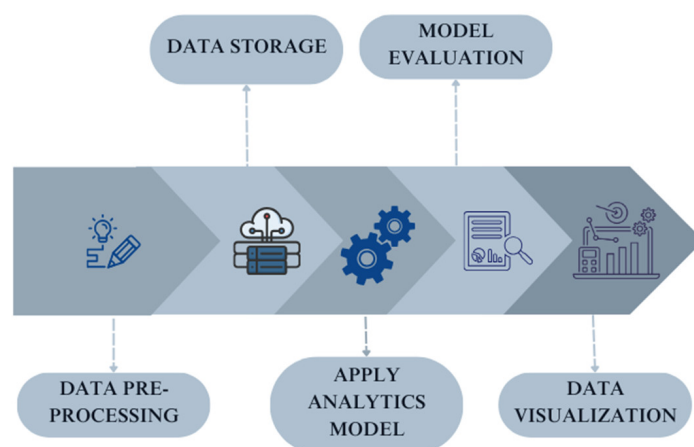
#### 4.3. Decision-Making Process, Decision-Maker, and Decision

Following the outcomes and analyses of the forecasting, the responsibility shifts to the decision-maker, in this case, the SETAs. The decision-maker is tasked with establishing a structured procedure or intervention protocol based on the insights gained from the forecasting model. It is crucial to note that while this work focuses on the first two elements, the final three are left to the discretion and expertise of the SETAs. This approach ensures a collaborative and adaptive decision-making process, aligning with the principles of DDDM and allowing SETAs to tailor interventions based on the unique demands of their sectors.

To enhance the presentation of forecast results, an interactive interface was developed using the Shiny package and R. This approach, rooted in data-driven decision-making (DDDM) principles, ensures an efficient and user-friendly platform for synthesizing outcomes to facilitate better decision-making. Within the Shiny interface, users can dynamically fine-tune ARIMA model parameters using interactive widgets, fostering adaptability and enabling a comprehensive exploration of the forecasting results. Chan et al. [57] emphasized that data visualization is a critical element in the DDDM framework. In DDDM, the effectiveness of the decision can either get better or suffer depending on the integrity of the data and the method used for analysis [52]. However, the decision quality is not only based on analysis, but also significantly affected by the data visualization [58]. This integration of technology not only aligns with DDDM principles but also elevates the user experience, promoting a more intuitive and informed interaction with the forecast results.

#### 4.4. Methodology Summative

The methodology, as illustrated in Figure 3, commenced with a crucial data handling phase, which involved data pre-processing aimed at refining extensive datasets to ensure data readiness and accuracy for subsequent analysis. This step involved meticulous cleaning, transformation, and structuring of occupational data, preparing it for integration into the Microsoft SQL database. Once curated, the data were loaded into the designated database, establishing a strong foundation for subsequent analytical steps. The subsequent analytical phase involved a pivotal connection between the SQL database and R, the programming tool employed in this study. This connection serves as the backbone for occupational analysis and the user interface created using the Shiny package. Employing R programming, the ARIMA model was applied to the data retrieved from the SQL server. To validate the model's accuracy, it was employed to forecast values already available, enabling performance evaluation. Following model evaluation, actual forecasting was executed. Finally, to streamline the decision-making process for SETAs, the analysis results were presented in a web platform using the Shiny package. Within the R Shiny environment, the interface was intuitively designed, equipping users with user-friendly tools for interactive exploration of datasets. Leveraging the R Shiny interactive features, users can effortlessly visualize intricate trends, delve into historical patterns, and engage in forecast analyses.



**Figure 3.** Methodology flow diagram.

## 5. Results

### 5.1. Pre-Processing

To ensure accurate forecasting, data pre-processing was essential to maintain data quality. The first step involved assessing missing values. Figure 4 (percentage of missing values) displays the proportion of missing values for each year in the dataset. The years 2016–2023 all showed relatively low percentages of missing data, with 2016 having ~3% of missing values. The remaining years exhibited low levels of missingness, indicating that the dataset was largely complete. Figure 4 (missing values in rows) further illustrates the missing data across rows. The plot illustrates that the missing data were scattered sporadically across different rows rather than being concentrated in specific areas. The sporadic pattern suggests that the missingness is likely to be Missing Completely at Random (MCAR).

This analysis indicated that the data were mostly complete, with only a small percentage of missing values. The sporadic and low-level missing data were likely MCAR, reducing the likelihood of bias in the predictions. However, to ensure accuracy and robustness, Multiple Imputation by Chained Equations (MICE) was employed, a method recommended to handle missing values [59]. This approach preserves statistical properties of the data by creating multiple imputations and combining them to form a complete

dataset. By using MICE, the small amount of missing data was effectively addressed, ensuring that the data remained robust for forecasting.



Figure 4. Missing values.

To assess the quality and consistency of the data from WSP and ATR, a linear mixed-effects model was employed, the result is illustrated in Table 2. This model is particularly suitable for the dataset, as it accounts for both fixed effects, such as year, and random effects, such as occupation codes, thereby capturing the inherent variability across different occupation codes over time [60]. The fixed effect of year had an estimated coefficient of 1.156, suggesting a slight upward trend in values over the years. However, this effect was not statistically significant ( $p$ -value = 0.365), indicating that the year-to-year variations were not substantial. The random intercept for occupation codes showed a variance of 409,995 with a standard deviation of 640.3, reflecting considerable variability among different occupation codes. Additionally, the residual variance was 39,653 with a standard deviation of 199.1, indicating the variability within each occupation code over time. The overall model fit, as indicated by the REML criterion, was satisfactory. The scaled residuals, which ranged from  $-14.7970$  to  $15.9067$ , mostly clustered around zero, further suggesting a good fit for most data points.

Table 2. Linear mixed effects.

Predictors	Value		
	Estimates	CI	$p$
(Intercept)	-2090.81	7142.37–2960.74	0.417
Random Effects			
$\sigma^2$	39,653.36		
$\tau_{00}$ occupation code	409,995.13		
N occupation code	595		
Observations	4673		
Marginal $R^2$ /Conditional $R^2$	0.000/0.912		



## 5.2. Trends

The advent of Industry 4.0 has had a huge impact on the manufacturing sector, not only in South Africa but around the world [12]. The impact of technology in the South African manufacturing sector is reflected in the employment trends within technology-centric and non-technology-centric roles, as illustrated in Tables 3 and 4. These tables combine employment data from the chemical and FoodBev sectors up to 2022, as this is the year for which FoodBev data were available. For the purposes of this analysis, ‘technology-centric’ jobs refer to jobs that necessitate some understanding of technology, encompassing software, digital systems, and specialized tools. These roles are fundamentally propelled by ongoing technological advancements, demanding expertise in areas such as computer science, programming, and digital infrastructure [19]. The trends in these tables provide an overview of how the South African manufacturing sector is responding to technological changes.

**Table 3.** Technology-centric jobs.

Occupation	2016	2017	2018	2019	2020	2021	2022
Data Management Manager	197	183	240	256	466	537	607
Information Technology Manager	109	119	87	107	127	170	156
Business Administrator	107	251	352	189	154	594	482
Software Architect	26	31	34	38	57	86	125
Engineering Planner	661	742	514	635	895	1079	1013
Data Capturer	190	192	129	196	188	212	206
Computer Analyst	169	187	172	159	221	492	499
Communications Analyst (Computers)	65	82	107	100	186	107	116

**Table 4.** Non-technology-centric jobs.

Occupation	2016	2017	2018	2019	2020	2021	2022
Procurement Administrator	800	740	772	820	595	575	594
Administration Clerk/Officer	3051	4707	3273	3073	2949	3126	2723
Call Center Customer Service Representative	203	295	195	240	54	52	29
Pay Clerk	208	190	186	177	168	183	174
Aisle Controller	1361	1629	1299	1307	1152	1090	1008
Delivery Clerk	2131	2833	2844	2453	1901	1982	1758
Manufacturing Store person	1868	1655	1848	1629	1398	1536	1749
Front-End-Loader Driver	488	108	101	474	230	197	136
Front Desk Coordinator	567	521	569	512	511	450	429
Regulatory Affairs Administrator	505	564	461	491	390	402	519

In line with prevalent literature [15,19,61] highlighting occupations such as software developers and ICT-related roles as pivotal in the era of Industry 4.0, a careful selection of these roles is presented as ‘technology-centric’ jobs in Table 3. Conversely, drawing from sources such as McKinsey [62], which underscore the transformation of roles involving physical labor and data collection in the advent of Industry 4.0, Table 4 features a curated compilation of these non-technology-centric jobs. This categorization aims to offer a distinct insight into the contrasting employment trajectories shaped by the influence of technology across diverse occupational domains.

The patterns within these technology-centric roles unveil a dynamic narrative of evolution and adaptation, directly correlated with the sweeping digitization across industries. Notably, roles such as Data Management Manager and Information Technology Manager experienced a pronounced surge in demand post-2019, signifying an accelerated growth phase within these areas. Contrasting this, the trajectory of Business Administrator positions showcased erratic fluctuations, possibly reflecting the changing demands and evolving responsibilities within administrative domains amidst digital transformation. Within this tech landscape, the ascent of Software Architect roles remained consistent,

albeit moderate, reflecting the steady but progressive nature of this specialized field. Engineering Planner positions, on the other hand, depicted a robust and consistent growth pattern, highlighting a substantial demand for specialized technical expertise. However, the trends observed in Data Capturer roles displayed more volatile fluctuations, indicating a potentially more sensitive response to digital advancements. Amidst these fluctuations, Computer Analyst roles exhibited a moderate trajectory, marked by occasional peaks, while Communications Analyst (Computers) positions portrayed a blend of stability and sporadic growth. Overall, these trends collectively depicted a landscape deeply influenced by the pervasive digitization across industries, emphasizing the critical need for evolving skill sets aligned with technological advancements, steering the course of these occupations' growth and prominence.

Looking at the occupations shown in Table 4, it is evident that these non-technological jobs experienced a downward trend in demand over the observed period. The observed decline in job roles such as Procurement Administrator/Coordinator/Officer, Administration Clerk/Officer, and Call Center Customer Service Representative (outbound) can be attributed to the transformative influence of technology on job functions. Advancements in automated procurement systems, streamlined administrative processes, and AI-driven customer service platforms have likely led to the diminishing need for these specific positions. The consistent downtrend in roles such as Pay Clerk indicates a shift toward more efficient payroll systems or automation in financial record-keeping. The decreasing demand for Aisle Controllers, Delivery Clerks, and Manufacturing Store Persons might be a consequence of supply chain innovations and automated warehousing systems, which reduce the requirement for manual inventory management and logistics handling. Similarly, roles such as Front-End-Loader Driver and Front Desk Coordinator may be impacted by technological advancements, such as self-service kiosks and digital reception systems, streamlining operations and minimizing manual involvement. The fluctuating yet declining trend in Regulatory Affairs Administrator positions may reflect digital regulatory platforms or more efficient compliance technologies, reducing the need for extensive manual regulatory oversight. Overall, these declining trends in job roles highlight the evolution of industries, with technology serving as a catalyst for optimizing processes, reducing manual tasks, and reshaping job demands. These observations are consistent with those of the authors of [18], who suggested that the advent of Industry 4.0 in South Africa will most likely decrease manual and repetitive jobs that can be easily automated. This is not only particular to the South African labor market, but it is observed all around the world. More specifically, it is anticipated that the advent of Industry 4.0 would result in the decline of low-skilled laborers and growth of high-skilled laborers [12].

Figure 3 illustrates the decline in low-skilled workers. The education level classification system according to the SETAs Skills Sector Plan (SSP) framework encompasses five distinct levels, as shown in Figure 5. Levels 0–1 represent individuals with no formal education or some basic schooling without a high school certificate, categorized as unskilled workers according to the South African Qualification Authority (SAQA). Individuals in levels 2–4 either have obtained their high school certificates or acquired basic training from Technical and Vocational Education and Training (TVET) institutes, earning national certificates: vocational (NCV) 1–3 certifications, and are considered semi-skilled. Level 5 represents individuals who have attained their high school certificates and pursued further tertiary education, ranging from NCV 4–6 to doctoral degrees, all categorized as skilled workers.

The labor force in South Africa is predominantly composed of semi-skilled workers, with a majority possessing NQF level 4, indicating that they either have attained their high school certificates or NCV-3 certificates. A significant decline in NQF level 0 (no schooling) was observed between 2019 and 2022, accompanied by an increase in NQF level 1 (some basic education), suggesting a positive trend toward skill enhancement in the labor force. The decline and increase were observed between levels 2 and 3, with level 2 decreasing while level 3 increased, suggesting some sort of skills enhancement in the manufacturing sector. A concerning trend regarding skilled workers at level 5 was observed. The observed

decline raises concern for the South African labor market, indicating potential challenges in retaining highly skilled individuals. Notably, the downward trend in the number of skilled workers was after 2020, which could be attributed to the implementation of remote work during the COVID-19 pandemic.

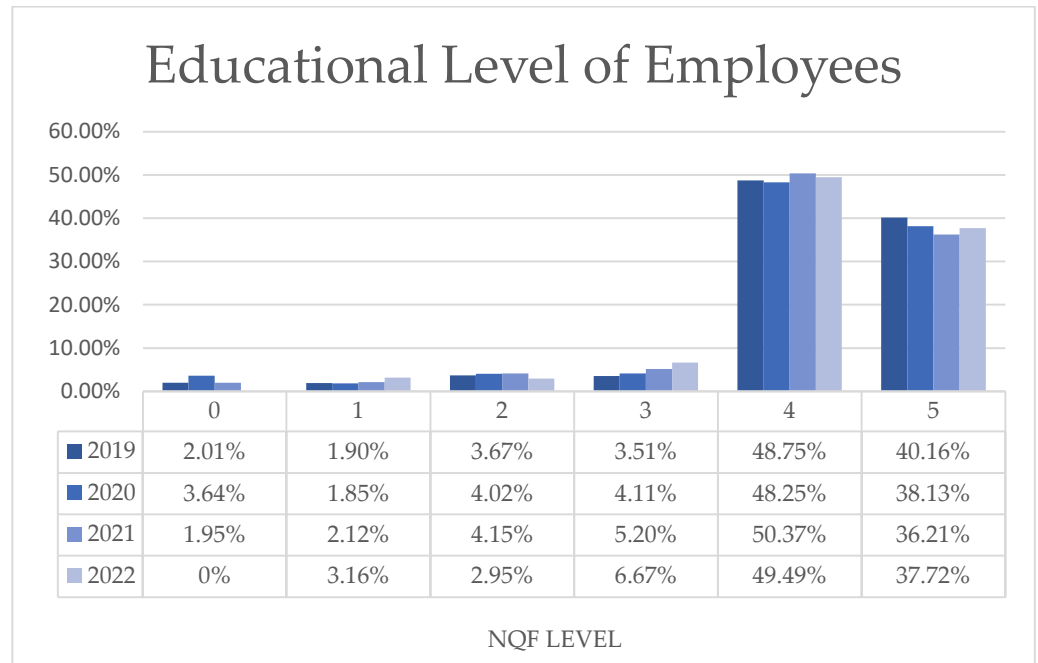


Figure 5. Educational levels of employees in the chemical sector 2019–2022.

Data visualization is a critical aspect of data-driven decision-making. Utilizing the Shiny package with the R programming language, an occupational analysis interface, as illustrated in Figure 6, was developed. This interface serves as an interactive platform for decision-makers to visualize the occupational trends in the sector. On the left side in the interface (indicated by the dotted box), there is an option to “select specialization”, allowing users to visualize trends for specific occupations. Figure 6 shows snippets of these occupational trends for different occupations. The interface features multiple tabs: a trend tab, correlations tab, and a forecast tab.

The occupational analysis interface stands as a ‘tangible’ manifestation of the ARIMA model applied in this work, delivering an interactive and research-oriented platform for time series analysis and forecasting. Rooted in the insights of [63], the interface emphasizes the importance of interactive web-based data visualization, ensuring a user-friendly and research-centric experience. This design choice resonates with the work of the authors of [64], who underscored the importance of graphical exploration in understanding time series data. The authors of [65] argued that effective time series visualization is crucial for uncovering patterns, and the chosen representation facilitates a clear examination of temporal trends. Furthermore, the emphasis on identifying patterns, such as growth or decline, within the trend tab echoes the advice in [66]. The authors stressed the importance of recognizing and interpreting patterns as a foundational step in decision-making. The trend tab, therefore, serves as a practical implementation of this advice, providing users with a clear and accessible tool to identify and comprehend temporal trends within the manufacturing sector.

# Chemical Sector Skills Forecast

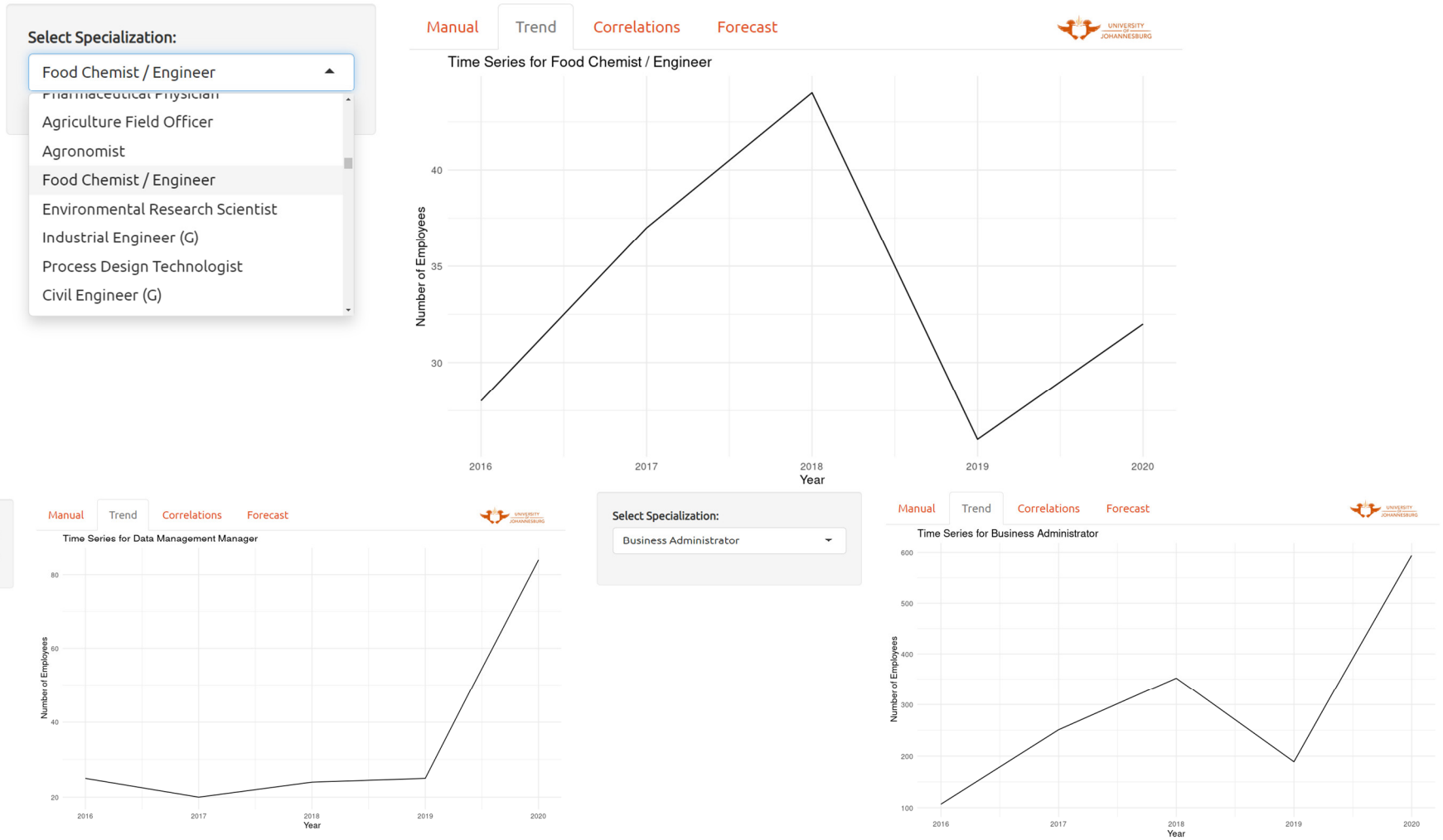


Figure 6. Cont.

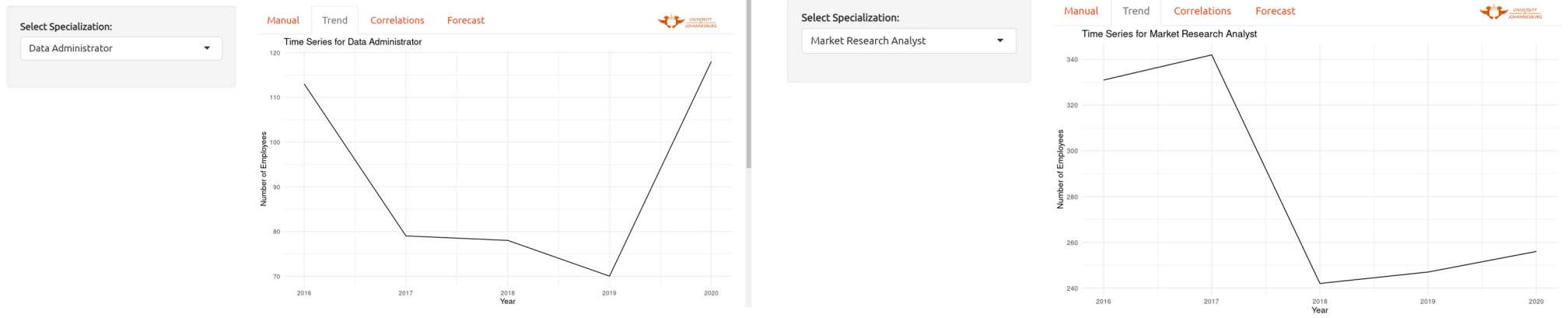


Figure 6. Occupational interface—trend tab.



### 5.3. Forecasts

The preceding section provided an overview of the employment trends in the manufacturing sector. The subsequent section shifts the focus toward predictive analytics using the ARIMA model. Using the historical employment data from the two sectors, projections one year into the future were conducted. The FoodBev and CHIETA datasets were forecasted separately due to different data lengths, with the FoodBev data spanning seven years (2016–2022) and the chemical sector spanning eight years (2016–2023). Before computing the actual projections, a preliminary forecast was performed using the available data, excluding the most recent year, to assess the accuracy of the model. Table 3 illustrates selected results for the chemical sector. The mean average percentage error (MAPE) was used to compare the predicted value and the actual value in order to evaluate the accuracy of the model.

The model's performance can be summarized as follows:

1. The FoodBev dataset consisted of 713 occupations, of which 473 (66%) were predicted with 80% and above accuracy.
2. The chemical sector consisted of 522 occupations, of which 474 (91%) were predicted with 80% and above accuracy.

The preliminary projections proved the model to be reliable. Furthermore, it is discernable that the model's performance improved with increased data availability, as indicated by the improved accuracy in the chemical sector projections, which had more data than the FoodBev sector. This outcome suggests the potential for enhanced accuracy with the inclusion of additional data for the excluded years in the preliminary projections. To enhance the decision-making process in terms of required interventions in response to the projected result, the final projections were only performed for occupations with a MAPE of no more than 20%, as shown in the MAPE column of Table 3. The threshold was implemented to ensure a high level of confidence in the accuracy and reliability of the projected results.

The forecasting results for the chemical sector, as shown in Table 5, provided insightful projections. Occupations with the most employees, such as Sales Representatives (Medical and Pharmaceuticals), Chemistry Technician, Chemical Plant Controller, and Chemical Engineering Technician, which are critical roles in the chemical sector, were projected to increase significantly in 2024. Some of the technology-centric roles, which were noted in Table 3, such as Database Manager and Database Designer and Administrator, were also expected to grow, and their projected increase was substantial.

The forecast tab within the occupational analysis interface is depicted in Figure 7. Complementing the trend tab showcased in Figure 6, this forecast tab stands as a pivotal component of the tangible artifact resulting from this work. By offering a visual representation of the forecasted results shown in Table 3, it empowers end-users to delve deeper into the intricacies of the projected occupational landscape. The quality of the decisions when using DDDM is largely affected by visualization [58]. This enhanced understanding facilitates informed decision-making, enabling stakeholders to discern and implement critical interventions in response to the forecasted demands. Ultimately, the occupational analysis interface serves as a cornerstone for proactive planning and strategic initiatives geared toward addressing the evolving needs of the manufacturing sector workforce.

The main objective of this study was to address the absence of forecasting methods within the South African manufacturing sector. In this work, a forecasting framework that produced reasonably reliable results was developed. Additionally, an occupational analysis interface was created, aligning with the principles of data-driven decision-making. The integration of this interface ensured that the forecasting framework was not only a theoretical model but also a practical tool that can significantly aid the manufacturing sector in addressing the skill demands.

**Table 5.** Forecast results for the chemical sector.

Occupation Title	2016	1017	2018	2019	2020	2021	2022	2023	Predicted 2023	MAPE	2024 Forecast
General Manager Public Service	17	6	6	40	19	23	63	46	40	13.32904	48
Trade Union Representative	29	37	48	37	19	38	31	28	26	8.013044	29
Human Resource Manager	358	310	341	344	421	453	434	452	388	14.1828	466
Business Training Manager	267	301	464	238	137	128	133	93	79	14.90708	96
Chief Information Officer	92	132	133	132	91	58	62	43	38	11.90166	46
ICT Project Manager	50	61	53	71	61	53	46	49	42	14.2755	50
Data Management Manager	25	20	24	25	70	84	83	76	64	15.46296	79
Financial Markets Business Manager	15	6	8	4	13	19	13	17	15	13.43848	18
Laboratory Manager	202	173	232	274	288	295	302	278	249	10.42141	284
Operations Manager (Non-Manufacturing)	106	154	179	81	178	238	210	183	157	14.32432	187
Importer or Exporter	67	46	58	54	47	53	65	39	35	10.43915	40
Retail Manager (General)	116	121	100	143	146	85	78	179	155	13.24364	184
Manufacture Research Chemist	50	79	87	72	112	100	70	119	103	7.355225	125
Retail Pharmacist	215	211	267	232	39	40	83	285	264	7.345125	298
Market Research Analyst	331	312	242	247	192	256	231	117	103	12.37104	123
Communication Coordinator	175	214	157	214	111	77	123	76	66	12.84006	78
Sales Representative—Medical and Pharmaceutical	2739	2668	2727	2595	2008	2620	2220	1838	1586	13.73677	1879
ICT Systems Analyst	164	179	172	159	221	492	499	493	411	16.57107	503
Database Designer and Administrator	110	67	78	70	107	118	548	126	110	12.42141	129
Librarian	14	20	33	36	15	15	17	15	13	13.0528	16
Information Services Manager	176	252	148	142	60	46	40	51	44	14.00231	54
Technical Director	24	208	28	32	19	17	24	22	19	11.84977	23
Chemistry Technician	2627	2737	2973	3552	2004	1996	2186	2306	1993	13.59102	2383
Radiation Control Technician	20	55	50	55	49	59	15	48	41	13.62571	51
Electrical Engineering Technician	293	405	536	551	264	286	398	486	420	13.59831	502
Mechanical Engineering Technician	665	522	434	451	483	587	730	768	665	13.42468	791
Pressure Equipment Inspector	110	58	73	74	96	82	115	92	78	15.21263	97
Chemical Engineering Technician	272	217	155	211	366	469	563	721	607	15.76788	760
Draughtsperson	141	169	248	192	172	174	162	153	132	13.60677	159
Water Plant Operator	28	46	72	29	39	36	37	77	67	12.89806	80
Chemical Plant Controller	5800	5818	5280	3458	5010	5290	5214	5347	4737	11.41503	5655
Gas or Petroleum Controller	1318	1164	1046	1637	607	627	907	523	450	14.05107	541
Manufacturing Production Technicians	66	95	275	247	669	768	468	442	399	9.622997	463

Table 5. Cont.

Occupation Title	2016	1017	2018	2019	2020	2021	2022	2023	Predicted 2023	MAPE	2024 Forecast
Health Technical Support Officer	4	13	3	2	31	43	52	19	16	13.17975	20
Sales Representative—Building and Plumbing Supply	82	3	36	61	49	100	50	60	50	16.22182	62
Sales Representative—Personal and Household Goods	191	384	197	380	314	680	489	157	138	12.38791	163
Commercial Services Sales Agent	14	8	19	15	25	11	31	51	44	12.84063	52
Manufacturer's Representative	28	10	10	14	57	5	15	40	34	14.67164	41
Chemical Sales Representative	1101	1091	937	874	751	667	822	955	816	14.54156	988
Property Manager	21	22	56	64	17	12	14	15	13	14.4805	16
Sales Representative—Business Services	137	493	361	440	476	134	374	228	208	8.785947	234
Waste Material Sorter and Classifier	1	3	12	8	6	2	2	99	84	14.70158	104
Handyperson	467	630	331	3333	1159	2299	719	571	485	14.97975	603
Chemical Mixer	291	195	310	211	1114	596	169	155	129	16.48581	158
Local Authority Manager	11	6	7	4	25	46	6	37	32	12.56638	38
Internal Audit Manager	23	18	15	19	38	22	26	31	27	11.98061	32
Recruitment Manager	10	13	15	19	11	9	9	12	10	13.03594	12
Quality Systems Manager	470	354	329	330	382	324	255	234	197	15.81662	242
Construction Site Manager	73	56	54	39	52	45	43	61	55	9.414381	62
Information Technology Manager	70	75	58	76	127	170	156	120	108	9.865707	123
Facilities Manager	104	109	104	89	89	94	85	206	183	11.03695	215
Electrical Specifications Writer	15	11	15	19	16	9	13	6	5	14.37255	6
Architect	1	6	11	4	9	2	8	7	6	13.74234	7

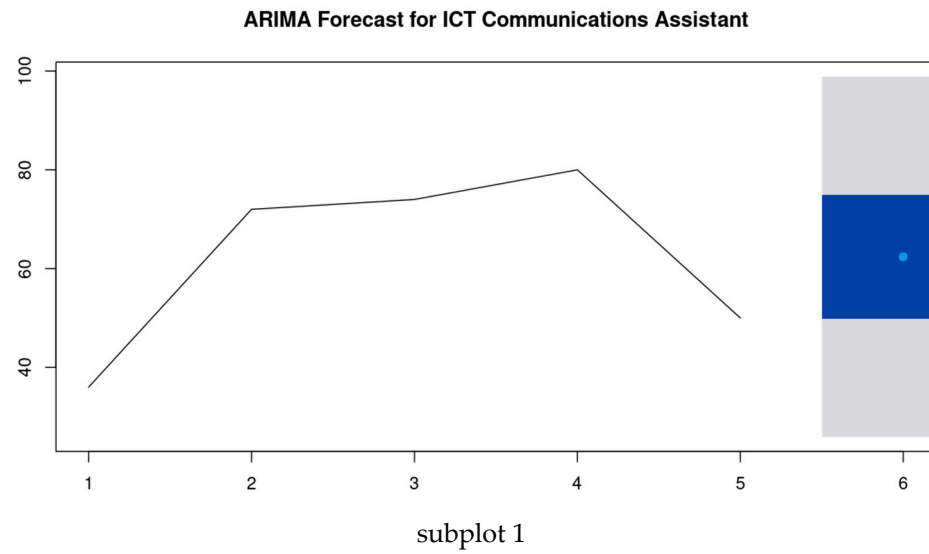
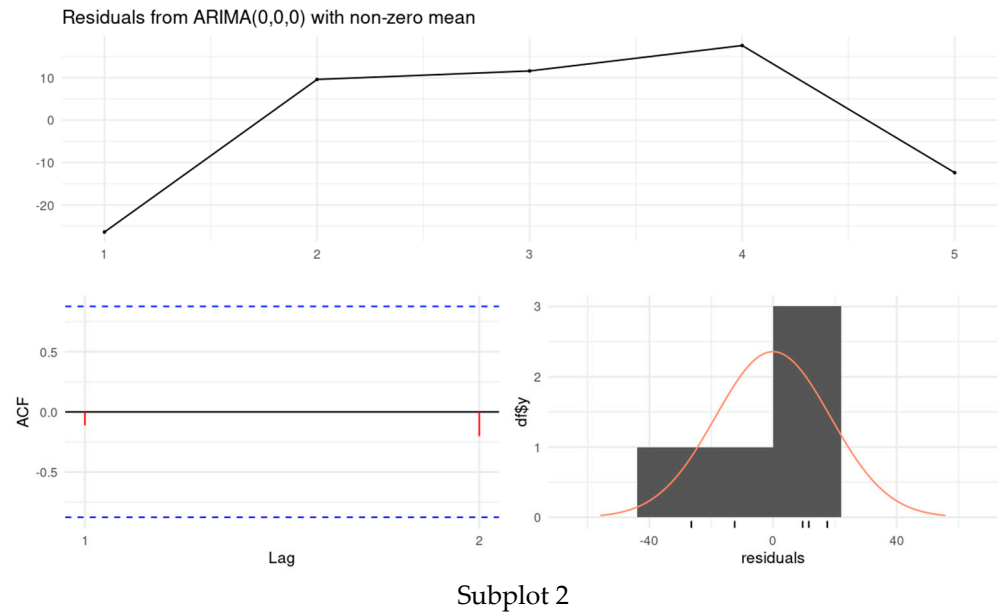
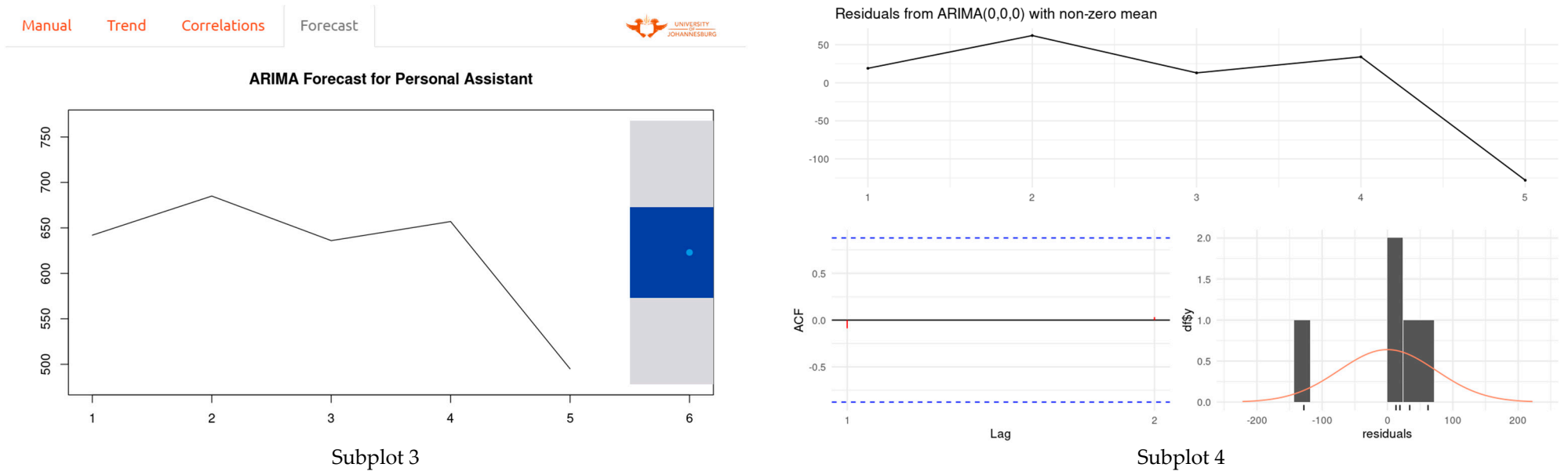


Figure 7. Cont.





**Figure 7.** Forecast tab.: The blue region in the plot represents the prediction interval boundaries for the forecasted values, specifically encompassing the 50% and 95% prediction intervals, with a central dot marking the point forecast, which is the ARIMA model’s estimate for the expected value. The ACF and PACF plots (blue dotted lines) illustrate the autocorrelation function and partial autocorrelation function of the residuals, which are crucial for diagnosing the fit of the ARIMA model by showing the correlation of residuals at different lags. The yellow curve (subplot 2 and 4) represents a normal distribution, which indicates how the residuals (the differences between observed and predicted values) are distributed around the zero mean.



## 6. Conclusions

As one of the cornerstones of the nation's economic growth, particularly in the context of Industry 4.0, the manufacturing sector faces transformative shifts. Technological advancements not only necessitate the acquisition of new skills but also signal changes in the profiles of existing occupations. Consequently, the need to anticipate and forecast future occupational demands within the manufacturing sector becomes increasingly evident. Through the application of the ARIMA model, a widely used tool in occupational forecasting, this study has provided valuable insights into the evolving employment landscape. By employing analytical tools, the research has paved the way for informed, data-driven decision-making and proactive skills planning for the manufacturing sector. The ARIMA model was applied to occupational data from the FoodBev and the chemical manufacturing sectors. The FoodBev dataset, spanning seven years (2016–2022), served to demonstrate the utility of the ARIMA model, while the eight-year (2016–2023) dataset from the chemical industry was utilized for the final projections. The validation step revealed a clear correlation between data volume and predictive accuracy, with the model accurately predicting 67% of the occupations with 80% accuracy in the shorter FoodBev dataset, and an impressive 91% of occupations with similar accuracy in the larger chemical sector dataset. These findings are in line with existing literature, such as [31], which emphasized the importance of data volume in enhancing predictive accuracy. The final projections focused solely on the chemical sector dataset, projecting occupational demand for the year 2024. The results indicated a notable increase in demand for traditional roles in the chemical sector, and a smaller demand for the technology-centric occupational profiles. This contrast suggests that the sector is not replacing the old with new, but gradually integrating new into the sector. Beyond the forecasting efforts, an occupational analysis interface was developed. This interface serves as a vital tool for end-users, providing them with a detailed and graphical view of the projected results. This is a global practice by sectoral bodies that conduct skills and occupational forecasts, such as BLS and CEDEFOP, who have online interactive graphics to view occupational trends. By offering enhanced visibility into the projected occupational demands, this interface empowers decision-makers to formulate targeted skills interventions and strategies, effectively aligning with the anticipated industry needs.

While this study has demonstrated the utility of the ARIMA model in labor market forecasts, several limitations must be acknowledged. Firstly, the data used were specific to FoodBev and CHIETA, potentially limiting the generalizability of the findings across other sectors. Additionally, while ARIMA is powerful, it has limitations in capturing seasonal effects [31,32]. Sectors with seasonal employment patterns, such as hospitality and tourism, may benefit more from the Seasonal ARIMA (SARIMA) model [56]. Lastly, the accuracy of the forecasts is strongly dependent on the data quality and volume, as evidenced by the contrast in accuracy for FoodBev and CHIETA, posing a challenge for sectors with less historical data.

## 7. Practical Implications

The study is significant from a practical perspective, as it addressed the current lack of forecasting models within the manufacturing sector. The developed model, demonstrating reasonable accuracy, along with the occupational interface can be adopted immediately to drive data-driven decision-making in the sector. With the accuracy presented for the CHIETA, the forecasting results have been incorporated into the Skills Sector Plan report for 2024. This alerts policymakers and stakeholders within the chemical sector who can leverage these insights for targeted skills' development and strategic workforce development, enhancing the sector's competitiveness and resilience in the future.

## 8. Future Work

The scope of this study can be expanded beyond the manufacturing sector to include other sectors as well. Future enhancements to the forecasting framework can incorporate factors such as technological advancements and economic influences on demand and

supply, similar to forecasting frameworks used in other developing countries. By doing so, the framework can be more robust and versatile, providing more comprehensive and accurate predictions across various sectors.

**Author Contributions:** Conceptualization, X.M., M.N. and A.T.; methodology, X.M., M.N. and A.T.; software, X.M. and M.N.; validation, X.M., M.N. and A.T.; formal analysis, X.M. and M.N.; investigation, X.M. and M.N.; resources, A.T.; data curation, X.M. and M.N.; writing—original draft preparation, X.M., M.N. and A.T.; writing—review and editing, X.M., M.N. and A.T.; visualization, X.M. and M.N.; supervision, A.T.; project administration, A.T.; funding acquisition, A.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research was funded by the CHIETA with previous research funded by the Foobev SETA.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data will be made available on request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Wong, J.; Chan, A.; Chiang, Y.H. A critical review of forecasting models to predict manpower demand. *Constr. Econ. Build.* **2004**, *4*, 43–56. [[CrossRef](#)]
2. Senthurvelautham, S.; Senanayake, N. A machine learning-based job forecasting and trend analysis system to predict future job markets using historical data. In Proceedings of the 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 7–9 April 2023; IEEE: New York, NY, USA, 2023; pp. 1–7.
3. Wilson, R.A.; Woolard, I.; Lee, D. *Developing a National Skills Forecasting Tool for South Africa*; Institute for Employment Research, University of Warwick: Coventry, UK; DoL/HSRC: Pretoria, Africa, 2004.
4. Arvan, M.; Fahimnia, B.; Reisi, M.; Siemsen, E. Integrating human judgement into quantitative forecasting methods: A review. *Omega* **2019**, *86*, 237–252. [[CrossRef](#)]
5. Flostrand, A.; Pitt, L.; Bridson, S. The delphi technique in forecasting—A 42-year bibliographic analysis (1975–2017). *Technol. Forecast. Soc. Chang.* **2020**, *150*, 119773. [[CrossRef](#)]
6. Ho, P.H.K. Labour and skill shortages in hong kong’s construction industry. *Eng. Constr. Archit. Manag.* **2016**, *23*, 533–550. [[CrossRef](#)]
7. Calonge, D.S.; Shah, M.A. Moocs, graduate skills gaps, and employability: A qualitative systematic review of the literature. *Int. Rev. Res. Open Distrib. Learn.* **2016**, *17*, 67–90. [[CrossRef](#)]
8. Corin, L. Job demands and job resources in human service managerial work an external assessment through work content analysis. *Old Site Nord. J. Work. Life Stud.* **2016**, *6*, 3–28. [[CrossRef](#)]
9. Whiteford, A.; Hall, E.J. *Sa Labour Market Trends and Future Work-Force Needs*; HSRC Bookshop: Pretoria, South Africa, 1999; pp. 1998–2003.
10. Woolard, I.; Kneebone, P.; Lee, D. Forecasting the demand for scarce skills, 2001–2006. *Hum. Resour. Dev. Rev.* **2003**, 458–474. Available online: <http://hdl.handle.net/20.500.11910/8092> (accessed on 30 June 2024).
11. Thomas, J. *Review of Best Practices in Labour Market Forecasting with an Application to the Canadian Aboriginal Population*; Technical Report; Centre for the Study of Living Standards: Ottawa, ON, USA, 2015.
12. Leit, P.; Geraldies, C.A.S.; Fernandes, F.P.; Badikyan, H. Analysis of the workforce skills for the factories of the future. In Proceedings of the 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS), Tampere, Finland, 10–12 June 2020; IEEE: New York, NY, USA, 2020; Volume 1, pp. 353–358.
13. Pinzone, M.; Fantini, P.; Perini, S.; Garavaglia, S.; Taisch, M.; Miragliotta, G. Jobs and skills in industry 4.0: An exploratory research. In *Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing, Proceedings of the IFIP WG 5.7 International Conference, APMS 2017, Hamburg, Germany, 3–7 September 2017*; Proceedings, Part I; Springer: Berlin/Heidelberg, Germany, 2017; pp. 282–288.
14. Jaschke, S. Mobile learning applications for technical vocational and engineering education: The use of competence snippets in laboratory courses and industry 4.0. In Proceedings of the 2014 International Conference on Interactive Collaborative Learning (ICL), Dubai, United Arab Emirates, 3–6 December 2014; IEEE: New York, NY, USA, 2014; pp. 605–608.
15. Prifti, L.; Knigge, M.; Kienegger, H.; Krmar, H. A competency model for “industrie 4.0” employees. In Proceedings of the *Wirtschaftsinformatik (WI) 2017*, St. Gallen, Switzerland, 12–15 February 2017.
16. Sackey, S.M.; Bester, A. Industrial engineering curriculum in industry 4.0 in a south african context. *S. Afr. J. Ind. Eng.* **2016**, *27*, 101–114. [[CrossRef](#)]
17. Chui, M.; Manyika, J.; Miremadi, M. *Where Machines Could Replace Humans—and Where They Can’t (Yet)*; McKinsey & Company: Chicago, IL, USA, 2016.

18. Maisiri, W.; Darwish, H.; Van Dyk, L. An investigation of industry 4.0 skills requirements. *S. Afr. J. Ind. Eng.* **2019**, *30*, 90–105. [[CrossRef](#)]
19. Frey, C.B.; Osborne, M.A. The future of employment: How susceptible are jobs to computerisation? *Technol. Forecast. Soc. Chang.* **2017**, *114*, 254–280. [[CrossRef](#)]
20. Charles, K.K.; Hurst, E.; Notowidigdo, M.J. Manufacturing decline, housing booms, and non-employment. *Chic. Booth Res. Pap.* **2013**, 13–57. [[CrossRef](#)]
21. Acemoglu, D.; Autor, D. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*; Elsevier: Amsterdam, The Netherlands, 2011; Volume 4, pp. 1043–1171.
22. Sandberg, J.; Holmstr, J.; Lyytinen, K. Digitization and phase transitions in platform organizing logics: Evidence from the process automation industry. *Manag. Inf. Syst. Q.* **2020**, *44*, 129–153. [[CrossRef](#)]
23. MacCrory, F.; Westerman, G.; Alhammadi, Y.; Brynjolfsson, E. *Racing with and Against the Machine: Changes in Occupational Skill Composition in an era of Rapid Technological Advance*; Association for Information Systems: Atlanta, GA, USA, 2014.
24. Rumberger, R.W.; Levin, H.M. Forecasting the impact of new technologies on the future job market. *Technol. Forecast. Soc. Chang.* **1985**, *27*, 399–417. [[CrossRef](#)]
25. Hughes, G. An overview of occupational forecasting in oecd countries. *Int. Contrib. Labour Stud.* **1994**, *4*, 129–144.
26. Bolli, T.; Zurlinden, M. Measurement of labour quality growth caused by unobservable characteristics. *Appl. Econ.* **2012**, *44*, 2297–2308. [[CrossRef](#)]
27. Garner, C.; Harper, J.; Howells, T.F., III; Russell, M.; Samuels, J. New bea-bls estimates of the industry-level sources of us economic growth between 1987 and 2016. *Int. Product. Monit.* **2019**, 187–203. Available online: <https://coilink.org/20.500.12592/7ht6h5> (accessed on 30 June 2024).
28. Khalaf, C.; Michaud, G.; Jolley, G.J. Predicting declining and growing occupations using supervised machine learning. *J. Comput. Soc. Sci.* **2023**, *6*, 757–780. [[CrossRef](#)]
29. Cedefop. *Future Skill Needs in Europe—Critical Labour Force Trends*; Publications Office: Luxembourg, 2016.
30. Cedefop. *Skills Forecast—Trends and Challenges to 2030*; Publications Office: Luxembourg, 2018.
31. Nwokike, C.C.; Okereke, E.W. Comparison of the performance of the sann, sarima and arima models for forecasting quarterly gdp of nigeria. *Asian Res. J. Math.* **2021**, *17*, 1–20. [[CrossRef](#)]
32. Sehwat, P.K.; Vishwakarma, D.K. Comparative analysis of time series models on COVID-19 predictions. In Proceedings of the 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 7–9 April 2022; IEEE: New York, NY, USA, 2022; pp. 710–715.
33. Noureen, S.; Atique, S.; Roy, V.; Bayne, S. A comparative forecasting analysis of arima model vs random forest algorithm for a case study of small-scale industrial load. *Int. Res. J. Eng. Technol.* **2019**, *6*, 1812–1821.
34. Rady, E.H.; Fawzy, H.; Fattah, A.M.A. Time series forecasting using tree based methods. *J. Stat. Appl. Probab.* **2021**, *10*, 229–244.
35. Zhang, D.; Chen, S.; Liwen, L.; Xia, Q. Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons. *IEEE Access* **2020**, *8*, 28197–28209. [[CrossRef](#)]
36. Rublikova, E.; Lubyova, M. Estimating arima-arch model rate of unemployment in slovakia. *Forecast. Pap. Progn. Pr.* **2013**, *5*, 275–289.
37. Weber, E.; Zika, G. Labour market forecasting in germany: Is disaggregation useful? *Appl. Econ.* **2016**, *48*, 2183–2198. [[CrossRef](#)]
38. Adenomon, M.O. Modelling and forecasting unemployment rates in nigeria using arima model. *FUW Trends Sci. Technol. J.* **2017**, *2*, 525–531.
39. Shobana, G.; Umamaheswari, K. Forecasting by machine learning techniques and econometrics: A review. In Proceedings of the 2021 6th International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 20–22 January 2021; IEEE: New York, NY, USA, 2021; pp. 1010–1016.
40. Elkamel, M.; Schleider, L.; Pasiliao, E.L.; Diabat, A.; Zheng, Q.P. Long-term electricity demand prediction via socioeconomic factors—A machine learning approach with florida as a case study. *Energies* **2020**, *13*, 3996. [[CrossRef](#)]
41. Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using cnn-lstm neural networks. *Energy* **2019**, *182*, 72–81. [[CrossRef](#)]
42. Muralitharan, K.; Sakthivel, R.; Vishnuvarthan, R. Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing* **2018**, *273*, 199–208. [[CrossRef](#)]
43. Dacrema, M.F.; Cremonesi, P.; Jannach, D. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In Proceedings of the 13th ACM Conference on Recommender Systems, Copenhagen, Denmark, 16–20 February 2019; pp. 101–109.
44. Mnyawami, Y.N.; Maziku, H.H.; Mushi, J.C. Comparative study of automl approach, conventional ensemble learning method, and knearest oracle-automl model for predicting student dropouts in sub-saharab african countries. *Appl. Artif. Intell.* **2022**, *36*, 2145632. [[CrossRef](#)]
45. Noureen, S.; Atique, S.; Roy, V.; Bayne, S. Analysis and application of seasonal arima model in energy demand forecasting: A case study of small scale agricultural load. In Proceedings of the 2019 IEEE 62nd International Midwest Symposium on Circuits and Systems (MWSCAS), Dallas, TX, USA, 4–7 August 2019; IEEE: New York, NY, USA, 2019; pp. 521–524.
46. Power, D.J.; Heavin, C.; Keenan, P. Decision systems redux. *J. Decis. Syst.* **2019**, *28*, 1–18. [[CrossRef](#)]

47. Elgendy, N.; Elragal, A.; Päiväranta, T. Decas: A modern data-driven decision theory for big data and analytics. *J. Decis. Syst.* **2022**, *31*, 337–373. [[CrossRef](#)]
48. Mandinach, E.B. A perfect time for data use: Using data-driven decision making to inform practice. *Educ. Psychol.* **2012**, *47*, 71–85. [[CrossRef](#)]
49. Kumar, A.; Shankar, R.; Aljohani, N.R. A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Ind. Mark. Manag.* **2020**, *90*, 493–507. [[CrossRef](#)]
50. Venkatramanan, S.; Lewis, B.; Chen, J.; Higdon, D.; Vullikanti, A.; Marathe, M. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics* **2018**, *22*, 43–49. [[CrossRef](#)]
51. Bourdeau, M.; Zhai, X.Q.; Nefzaoui, E.; Guo, X.; Chatellier, P. Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustain. Cities Soc.* **2019**, *48*, 101533. [[CrossRef](#)]
52. Janssen, M.; Van Der Voort, H.; Wahyudi, A. Factors influencing big data decision-making quality. *J. Bus. Res.* **2017**, *70*, 338–345. [[CrossRef](#)]
53. Elgendy, N.; Elragal, A. Big data analytics: A literature review paper. In *Advances in Data Mining. Applications and Theoretical Aspects, Proceedings of the 14th Industrial Conference, ICDM 2014, St. Petersburg, Russia, 16–20 July 2014*; Proceedings 14; Springer: Berlin/Heidelberg, Germany, 2014; pp. 214–227.
54. Theodosiou, M. Forecasting monthly and quarterly time series using stl decomposition. *Int. J. Forecast.* **2011**, *27*, 1178–1195. [[CrossRef](#)]
55. Amar, S.; Sudiarso, A.; Herliansyah, M.K. The accuracy measurement of stock price numerical prediction. *J. Phys. Conf. Ser.* **2020**, *1569*, 032027. [[CrossRef](#)]
56. Frechtling, D. *Forecasting Tourism Demand*; Routledge: London, UK, 2012.
57. Chan, K.-S.; Ripley, B.; Chan, M.K.-S.; Chan, S. *Package ‘tsa’*. *R Package*; Version 1; The Comprehensive R Archive Network (CRAN): Vienna, Austria, 2022.
58. Svensson, R.B.; Feldt, R.; Torkar, R. The unfulfilled potential of data-driven decision making in agile software development. In *Agile Processes in Software Engineering and Extreme Programming, Proceedings of the 20th International Conference, XP 2019, Montreal, QC, Canada, 21–25 May 2019*; Proceedings 20; Springer: Berlin/Heidelberg, Germany, 2019; pp. 69–85.
59. Van Buuren, S.; Groothuis-Oudshoorn, K. mice: Multivariate imputation by chained equations in r. *J. Stat. Softw.* **2011**, *45*, 1–67. [[CrossRef](#)]
60. Pinheiro, J.C.; Mbates, D. Linear Mixed-Effects Models: Basic Concepts and Examples. In *Mixed-Effects Models in S and S-Plus*; Springer: Berlin/Heidelberg, Germany, 2000; pp. 3–56.
61. Kvetan, V.; Wilson, R.; Zunkersteinova, A. Cedefop’s skills supply and demand forecast: 2011 update and reflections on the approach. In *Building on Skills Forecasts—Comparing Methods and Applications*; Publications Office of the European Union: Luxembourg, 2021; p. 11.
62. Woetzel, J.; Madgavkar, A.; Gupta, S. *India’s Labour Market: A New Emphasis on Gainful Employment*; McKinsey Report; McKinsey & Company: Chicago, IL, USA, 2017.
63. Sievert, C. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*; CRC Press: Boca Raton, FL, USA, 2020.
64. Hyndman, R.J.; Athanasopoulos, G. *Forecasting: Principles and Practice*; OTexts: Melbourne, Australia, 2018.
65. Cleveland, W.S. *Visualizing Data*; Hobart Press: Troy, OH, USA, 1993.
66. Shumway, R.H.; Stoffer, D.S. *Time Series Analysis and Its Applications: With R Examples*; Springer: Berlin/Heidelberg, Germany, 2006.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.