CJBEHM
CANADIAN JOURNAL OF BUSINESS, ECONOMICS AND HEALTH MANAGEMENT

ORIGINAL ARTICLE

CJBEHM. 2026 April; 2(1): 01-32
Published by www.cjbehm.ca

Time Series Analysis of Canadian Power Engineer and Power System Operator Occupational Vacancies and Vacancy Predictors

Clayton Roy Mullen, PhD

1. UBC Senior Thermal Energy Manager

Corresponding author*

Clayton Roy Mullen, PhD

UBC Senior Thermal Energy Manager

Email: crmullen@telus.net

Article info

Received: 24 January 2026
Accepted: 29 July 2026
Published: 30 April 2026

Keywords

Canadian power engineering, national occupational classification (NOC), time series, machine learning, ARIMA, SARIMA, LSTM

Abstract

This quantitative and exploratory investigation examines prospective trends in the job vacancies within the National Occupational Classification (NOC) 92100 using autoregressive integrated moving-average (ARIMA), seasonal autoregressive integrated moving-average (SARIMA), and long short-term memory (LSTM) univariate and multivariate models, along with their primary predictors. NOC 92100 encompasses both Canadian power engineers and power system operator classifications. The study aims to delineate mechanisms for future research on NOC 92100 by utilizing available Statistics Canada secondary data, with a particular emphasis on power engineering. The focus is on methodological insights rather than on specific numerical findings, thereby providing guidance for subsequent research and career development considerations. Analysis of vacancy data for NOC 92100 from 2015 to 2024 revealed a marked increase during the COVID-19 pandemic, peaking at 830 vacancies. Forecasting models, including ARIMA, SARIMA, and LSTM, project a more stable vacancy count of 600-700 from 2024 to 2027. The multivariate LSTM model demonstrated superior performance relative to ARIMA, SARIMA, and univariate LSTM models, as evidenced by metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Symmetric Mean Absolute Percentage Error (SMAPE). Correlation analyses indicated a significant association between NOC 92100 vacancies and Real Household Economic Activity (Factor 2), with Pearson's correlation coefficient $r = 0.776$, $p < 0.05$, and Spearman's rho = 0.849, $p < 0.05$. These findings are further substantiated by a feature ablation test, which yielded a Factor 2 ΔRMSE of 149.06.

Copyright: © by the authors. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) International license.

Introduction

Canadian power engineering offers a promising field for exploration, given the limited but growing amount of formal research specific to the discipline. Seminal studies in power engineering lay a foundational basis for further investigation and ongoing research, aligning with existing knowledge, research practices, and normal science (Kuhn, 1997). Research in this area has focused on education and training, workforce shortages, attracting and retaining talent, and the certification process (Boskovic, 2017; Ellis, 2010; Hau, 2007; Lutchman, 2008; Mullen, 2017). A knowledge gap remains regarding the future of power engineer and power system operator job vacancies, and the factors that influence them. Locating comprehensive secondary datasets specific to power engineering that are suitable for analysis is challenging. Various open-access repositories were examined to support the collection, aggregation, and analysis of relevant secondary data. These sources included Statistics Canada, Wharton Research Data Services (WRDS) (Wharton University of Pennsylvania, 2026), and relevant open-source databases. Statistics Canada was the most suitable data source for this study. Statistics Canada started collecting data on power engineering and power system operator vacancies in 2015. NOC information is gathered through the Statistics Canada Labor Force Survey (LFS) and the Job Vacancy

and Wages Survey (JVWS) (Statistics Canada, *LFS*, 2025; Statistics Canada, *JVWS*, 2025). The rapid development of data science, artificial intelligence (AI), and machine learning enables more detailed research exploration (Alharbi & Al-Alawi, 2024; Castaño Camps, 2022; Glackin & Adivar, 2023; Goodfellow et al., 2016). Statistics Canada consolidates power engineer and power system operator vacancy data into a single NOC 92100 category. As a result, vacancy data for power engineering cannot be reliably distinguished from that of power system operators. Both disciplines are included within the same NOC 92100 category in this study. Data for the selected exogenous factors, such as labor force characteristics, employment, interest rates, consumer price index (CPI), inflation, household income, and gross domestic product (GDP) that may potentially influence NOC 92100 were also obtained from Statistics Canada. The primary goal of this exploratory research is to provide a theoretical forecast of occupational vacancies for NOC 92100, identify broader exogenous factors that may affect them, and lay the groundwork for future research. The study analyzes the integrated NOC 92100 data and primarily focuses on its implications for the power engineering occupation.

The dataset used in this study included all available NOC 92100 vacancy data from Statistics Canada's database, covering 2015 to 2024, along with selected exogenous variables that may influence NOC 92100 occupational vacancies. This exploratory research compares the effectiveness of statistical time-series methods and machine learning approaches for analyzing NOC 92100 data and potential factors affecting NOC 92100. This investigation aims to add to existing knowledge and inform future research on the projected job outlook for the Canadian power engineering component of NOC 92100. Although progress has been made in this area, additional research is needed to accurately estimate future NOC 92100 vacancies and to identify the general factors influencing these occupational trends. Statistics Canada is the primary national agency that provides Canadians with essential data on the country's workforce, economy, environment, and related characteristics, using digital tools such as the LFS (Lalé, 2024; Statistics Canada, *LFS*, 2025). The most comprehensive and reliable Canada-wide data on power engineer and power system operator vacancies, in terms of quality and volume, is stored in the Statistics Canada database under NOC 92100. According to Statistics Canada criteria, an occupation (job) is considered vacant if it meets these conditions: (a) a specific position exists; (b) work could start within 30 days; and (c) the employer is actively seeking workers from outside the organization to fill the position (Statistics Canada, *Section 3: Terms and definitions*, 2025). The year-month-job category unit of analysis includes both power engineers and power system operators, classified under Statistics Canada's National Occupational Classification (NOC) codes 92100 and 9241. The 92100 NOC data used in this study corresponds to the 2021 census, while the 9241 NOC pertains to the 2016 census (Statistics Canada, *NOC*, 2025). Statistics Canada's National Occupational Classification (NOC) and the North American Industry Classification System (NAICS) provide comprehensive frameworks for classifying occupations and industries. These systems use both worker and industry data to describe employment supply (workers) and demand (jobs) within specific sectors (Statistics Canada, *Standards, data sources and methods*, 2025). The relationship between NOC and NAICS codes helps reconcile specific occupations within industries. The NOC focuses on a worker-centered, bottom-up approach that emphasizes occupational traits on the supply side. NAICS employs an industry-

centered, top-down approach highlighting employment needs from the demand side (Statistics Canada, *North American Industry Classification System,* 2024). The link between NOC and NAICS illustrates the relationship between labor market supply and demand. NAICS classifies industries, while NOC classifies occupations (Statistics Canada, *Standards, data sources and methods,* 2025). A workplace typically employs individuals with various job roles. These roles are grouped into specific occupational categories (NOCs). However, the same industry classification (NAICS) applies to all employees at a business or organization, based on the type of production process.

Power engineers are licensed to operate and maintain equipment such as pressure vessels, boilers, turbines, generators, reactors, engines, and auxiliary systems to supply utilities like heat, electricity, refrigeration, and lighting for industrial, institutional, and commercial users (B.C. Reg. 104/2004 M62/2004, 2026; Government of Canada, 2025). Power engineers receive academic and practical training in areas including thermodynamics, mechanics, electricity, chemistry, hydrodynamics, operations, and leadership. Power system operators oversee electrical equipment in control centers that manage the distribution and transmission of power across utilities, power plants, and various sectors (Government of Canada, 2025). The 92100 NOC code encompasses these roles and was among the 824 classifications available through Statistics Canada at the time of this study (Statistics Canada, Standards, data sources and methods, 2025). Research related to this NOC code offers insights into labor market trends, vacancy patterns, and strategies for addressing shortages or surpluses, which are essential for understanding current and future vacancy counts and their influencing factors. The Canadian power engineering hierarchical certification system varies by province and generally consists of four or five levels, from entry-level to managerial positions. Certification begins at the 4th or 5th class level and may advance to the 1st class. Alberta, British Columbia, Manitoba, and Saskatchewan have five levels, while other provinces and territories have four (Government of Canada, 2025). Statistics Canada does not track power engineering data by specific certification levels, which limits analysis at each level. Each certification level allows power engineers to handle increasingly complex systems and expands their job opportunities (ABSA, 2026; BCIT, 2026; TSBC, 2026). Engineers can progress from 5th to 1st class or remain at their current level. Certification enables them to operate or supervise equipment within regulatory limits, based on plant capacity, at their certified level and subordinate levels (TSBC, 2026). Consultations with key groups, such as the Institute of Power Engineers (IPE), provided insights and highlighted areas for intervention.

Literature Review and Theoretical Framework

The literature review examines time series analysis methods, including ARIMA, SARIMA, and LSTM models. The review addresses broader potential labor-market influences, including wealth, household economics, and the impact of technological advances and automation on human capital, all of which are interconnected with the macroeconomic factors. Studies by Sirisha et al. (2022), Selmy et al. (2024), Gopu et al. (2023), and Ullah et al. (2024) collectively underscore the evolving landscape of time series forecasting, particularly when comparing traditional stochastic models with advanced machine learning techniques such as LSTMs and their variants. Sirisha et al. (2022) demonstrated that LSTM surpassed both ARIMA and SARIMA in forecasting profit values derived from sales data spanning over four decades. Despite a marginal increase in accuracy (~3%) over SARIMA, LSTM's higher computational cost and

sensitivity to epoch tuning were noted. Similarly, Selmy et al. (2024) evaluated time-series models for IoT-based temperature prediction and introduced a hybrid CNN-LSTM model. The results showed that deep learning models outperformed traditional stochastic methods, with CNN-LSTM slightly outperforming LSTM across RMSE, MAE, and MAPE. Gopu et al. (2023) extended this evaluation to multivariate time series modeling for forest fire prediction in Portugal. Using weather and land-use data, the study confirmed the effectiveness of ARIMA, SARIMA, LSTM, and GRU, with LSTM being the most accurate. Ullah et al. (2024) validated LSTM's forecasting ability for dengue fever cases using a multivariate dataset including climate and epidemiological variables. LSTM outperformed GRU and RNN, achieving 87.98% accuracy, highlighting its usefulness in health forecasting domains involving complex temporal and variable interactions.

Understanding the attributes of ARIMA, SARIMA, and LSTM helped select the most suitable algorithm for NOC 92100 data. The literature consistently ranks LSTM as one of the top time series predictors, depending on the size and nature of the dataset (Gajewski et al., 2023; Gopu et al., 2023; Ibrahim et al., 2024; Kontopoulou et al., 2023; Rahhal et al., 2024; Selmy et al., 2024; Sirisha et al., 2022; Ullah et al., 2024). Factors such as wealth, GDP, household economic activity, inflation, and unemployment can influence the labor market and occupational fields, such as NOC 92100. These factors are influenced by broader trends such as technological change and automation. Gil-Alana et al. (2025) employed a time-series Autoregressive Fractionally Integrated Moving Average (ARFIMA) model to examine labor market trends in unemployment and job vacancies across the G7 nations (Canada, France, Germany, Italy, Japan, the UK, and the USA) from January 2002 to October 2023. A Beveridge Curve analysis revealed an increase in job vacancies following the COVID-19 pandemic (March 2020 to May 2023) and the December 2007 to June 2009 Great Recession across all G7 countries, except the UK. Okun's law describes the relationship between unemployment and economic output (GDP) through the Okun coefficient, which indicates that higher employment correlates with higher GDP. It predicts that a 1% drop in employment leads to a 2% decline in GDP (Cárdenas, 2024; Mohamed, 2024). Coşar & Yavuz (2021) examined the links between unemployment, changes in GDP, and labor force participation in Turkey from 1989 to 2019, noting that Okun's law coefficients rose during recessions. Mavrodi et al. (2024) studied 257 European NUTS2 regions post-2009 crisis through the pre-COVID period. They found that regional differences in the Okun coefficient stem from labor-market structures such as skill levels, unemployment duration, industry shifts, and business cycles.

Cardona-Arenas (2025) assessed the persistence of high unemployment in Organization for Economic Co-operation and Development (OECD) member countries, using labor income share in Gross Domestic Product (GDP), productivity, income inequality, and human capital accumulation from 2013 to 2019. The study found that countries with a higher labor income share of GDP, sufficient productivity to support well-being, and a higher human capital index had a lower probability of high unemployment persistence. Lin et al. (2023) found that unemployment and inflation rates mirror fluctuations in the capital and labor markets. Unemployment rates serve as an indicator of labor market flexibility, where inflation rates influence capital acquisition costs and subsequent capital allocation. The Bank of Canada (BoC) CPIX measure of core inflation excludes the eight most volatile components (Sun & Dimiski, 2024). This core inflation measure provides a more precise gauge of economic conditions than total CPI inflation and was included as a potential predictor in this study. Chemlal and Benomar (2024) employed a machine–learning–based time-series model to predict the impact of technological innovation on occupational employment demand in Spain. Analyzing Labor Force

Survey (LFS) data from 2011 to 2022, grouped by the National Classification of Occupations (CNO), the study forecasted demand through 2035. Results indicate that automation will influence no more than 5% of occupations, with 3–5% of both high- and low-skilled workers affected. Low-skilled workers engaged in manual and routine tasks are more exposed in absolute numbers. Automation primarily alters the job content rather than eliminating roles (Chemlal & Benomar, 2024; Marinaș et al., 2024). Becker's Human Capital Theory (HCT) (Becker, 1962; Weiss, 2015) explains how automation impacts employment, emphasizing education and skills development to mitigate these effects. HCT links training and education to economic productivity (Auerbach & Green, 2024), viewing human capital as a resource that enhances labor productivity, requiring strategic career management. Governance at societal and labor-market levels is vital to keep technological risks, such as automation, within acceptable limits (Bengio et al., 2024). Critics argue that increased education does not always lead to higher productivity or wages (Auerbach & Green, 2024). Krstic (2024) found that AI's effects on employment, skills, wages, and economic growth are complex and often contradictory, producing both innovation-driven gains and job losses in automatable sectors, alongside advancements in more sophisticated industries (Romer, 1990; Schumpeter, 2021; Solow, 1957). The 'jobless future' concept (Hanna et al., 2024), driven by automation, highlights the dual perspectives of technology replacing human workers and creating new job opportunities. This aligns with Schumpeter's 'creative destruction' theory (Schumpeter, 2021), which posits that job losses can lead to new employment as technologies displace older industries, fostering economic growth. AI is seen as both rendering some jobs obsolete and generating new industries. The spectrum of optimism and pessimism is reflected in labor-saving versus labor-supportive innovations (Sasaki, 2024). Bengio et al. (2024) addressed the imbalance between AI's benefits and harms, raising the Alignment Problem that considers whether AI aligns with human goals and remains controllable (Arbel et al., 2024; Bruiger, 2025). While AI advances medicine, ecosystems, and living standards, risks include malicious use, societal harm, and loss of human oversight. Prorok and Takacs (2024) emphasized AI's potential in corporate decision-making but cautioned that governance challenges and the need for self-correcting systems could hinder these benefits. Khalaf et al. (2023) evaluated machine learning techniques predicting occupational trends, finding that physical, repetitive tasks are key indicators of decline, while roles involving information and communication are linked to growth. As society favors interpersonal and digital careers, physical jobs are diminishing.

Artificial Intelligence and Machine Learning

Computers recognize and manage tasks that follow formal mathematical rules through programmed knowledge. AI uses computing to solve problems that may be difficult or intractable for humans. A goal of AI is to imitate or surpass human task performance (Agrawal et al., 2019; Townsend et al, 2024). AI reduces uncertainty by developing machine-intelligent systems with advanced predictive abilities. Machine learning, a subset of AI, uses data and algorithms, such as neural networks, to mimic human learning and improve accuracy (An et al., 2024; Castaño Camps, 2022; Goodfellow et al., 2016). Neural networks enable computers to learn by identifying and extracting patterns from raw data (Agrawal et al., 2019; Box et al., 2015; Goodfellow et al., 2016).

Time Series

Time series analysis is a statistical and machine learning approach used to predict target values from historical time-ordered data (Box et al., 2015; Sirisha et al., 2022). It aims to summarize and forecast temporal patterns by analyzing data aggregated at consistent

intervals. Analytical methods identify trends and recurring patterns, while predictive models extrapolate future values from past observations (Gopu et al., 2023). Accurate modeling requires continuous, regularly spaced data, with missing values addressed through replacement or imputation (Beard et al., 2019; Jin, 2024). Traditional models such as ARIMA and SARIMA perform well on small, cyclical datasets, whereas LSTM networks are better suited for large, complex, and incomplete time series (Cheng et al., 2023; Ullah et al., 2024).

ARIMA – Autoregressive Integrated Moving Average

ARIMA, introduced by Box and Jenkins in the 1970s (Box et al., 2015), is a linear stochastic model for univariate time series forecasting. It predicts future values using past observations, including lags and lagged errors (Cheng et al., 2023; Sirisha et al., 2022), and is well suited to data with stable trends (Huruta, 2024; Kontopoulou et al., 2023; Yamak et al., 2019). The model assumes linear relationships, is sensitive to outliers (Ibrahim et al., 2024), and requires non-stationary data to be differenced to achieve stationarity (Siami-Namini et al., 2018). ARIMA consists of AutoRegressive (AR), Integrated (I), and Moving Average (MA) components, each defined by non-negative integer hyperparameters (p, d, q). The parameter 'p' represents the order of autoregression, '**d'** is the number of differencing steps for stationarity, and '**q' is** the moving average order that models past error effects (Cheng et al., 2023; Kapadia & Mehta, 2025; Kontopoulou et al., 2023). ARIMA models are computationally efficient and effective for short-term forecasts of linear data, but they struggle with non-stationary or nonlinear patterns. They remain a helpful benchmark for advanced methods such as LSTM, which can capture complex relationships and reduce error through iterative optimization (Ibrahim et al., 2024; Zeng et al., 2023). ARIMA applies to unemployment and vacancy predictions (Gajewski et al., 2023) and can be extended with SARIMA to incorporate seasonality if necessary.

SARIMA – Seasonal Autoregressive Integrated Moving Average

The SARIMA model extends ARIMA by explicitly modeling applicable seasonal patterns, making it suitable for time series with both trends and seasonality. It uses seasonal differencing and parameters [P, D, Q, m] to handle complex periodic fluctuations (Byju et al., 2024; Huruta, 2024; Kontopoulou et al., 2023; Jain et al., 2024). While differencing helps achieve stationarity, it can also decrease accuracy. For nonlinear or long-term dependencies, LSTM networks often outperform ARIMA and SARIMA (Cheng et al., 2023; Sirisha et al., 2022). Symbolic AI (SAI) employs rule-based logic and explicit symbols for transparent reasoning but lacks adaptability because it relies on rigid knowledge bases. Connectionist AI (CAI), inspired by neural networks, learns from data and can overcome this rigidity, but it often operates as a "black box" (An et al., 2024). LSTMs, a type of recurrent neural network (RNN), exemplify the CAI approach by learning temporal patterns directly from data rather than predefined symbolic rules.

LSTM

LSTM, a deep learning architecture within the artificial neural network (ANN) framework, provides a machine learning alternative to traditional statistical forecasting models, such as ARIMA and SARIMA. Its adaptability, multivariate capacity, and ability to model nonlinear relationships often enable superior performance on large, complex datasets (Cheng et al., 2023; Kontopoulou et al., 2023; Yamak et al., 2019). However, small datasets may limit the effectiveness of LSTMs due to insufficient learning data. LSTMs employ activation functions, including the sigmoid for gate control and the hyperbolic

tangent (tanh) for the memory cell, to capture intricate temporal patterns (Cheng et al., 2023). LSTMs are effective for modeling nonlinear patterns in time series data and are not constrained by strict stationarity assumptions, enabling broader applications in real-world settings (Cheng et al., 2023). LSTMs can identify long-term dependencies, which are difficult for ARIMA-type models. The LSTM network is an improved version of the RNN architecture that distinguishes itself from RNNs by its ability to retain information over extended periods, thanks to its long-term dependencies and recursive structure (Johnston et al., 2025; Ullah et al., 2024; Waqas & Humphries, 2024). LSTM advantages include the ability to handle long-term dependencies by preventing vanishing gradients (Johnston et al., 2025; Waqas & Humphries, 2024), suitability for both univariate and multivariate data, inherent scalability, and LSTMs broad applicability to various sequencing tasks (Jin, 2024). Disadvantages include the need for larger datasets and greater computational power, lower interpretability compared to ARIMA or SARIMA, and difficulty in tuning and validating. LSTMs have inherent limitations despite their clear benefits over traditional time series models. These include the need for sufficient data to train the model, significant computational resources, and the 'black box' nature of LSTM (Ullah et al., 2024; Waqas & Humphries, 2024).

Methodology

The research method comprises two scenario-based phases that employ statistical, machine-learning, and correlational analyses. The first phase employs a trend scenario to predict future occupational vacancies for the NOC 92100 domain. This phase uses ARIMA, SARIMA, and LSTM univariate models. The univariate data is sourced from Statistics Canada's quarterly vacancy reports for NOC 92100, spanning 2015 to 2024. The second phase features an impact scenario that predicts future occupational vacancies in the NOC 92100 domain using a multivariate time-series LSTM model with exogenous factors. Subsequently, an assessment is performed to determine which factors or predictors most significantly affect NOC 92100 occupational vacancies, using Pearson's r and Spearman's rho correlation analyses, along with additional feature ablation and Steiger's Z test to evaluate the statistical significance of differences in dependent correlations. The secondary data for the NOC 92100 study were obtained from the Statistics Canada website as part of a larger dataset with 2,197 rows and 40 columns. A subset was extracted, consisting of 122 observations (rows) and 30 measures (columns), including date and job vacancy counts, spanning 2015 to 2024. The dataset includes the reference date and NOC 92100 occupational vacancy counts, along with related measures including labor force characteristics, employment by industry, interest rates, consumer price index (CPI), inflation, household income, GDP income, and GDP NAICS. Selecting the appropriate variables for a machine-learning model is crucial to ensure the model performs accurately and yields meaningful results (Box et al., 2015; Goodfellow et al., 2016; Orozco et al., 2024; Suman et al., 2024). The study follows the standard 6-stage Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which includes: (a) understanding the business for problem framing and goal setting; (b) understanding the data to collect and explore available information; (c) preparing the data for analysis; (d) modeling to choose the correct algorithm and develop the model; (e) evaluating to identify the best-performing model; and (f) deploying the model in a real-world setting (Getachew et al., 2024). Python and the Colab environment were selected for this study. Data preprocessing involved setting the date as the index, examining and managing missing values, and outliers. This was followed by exploratory data analysis to gain insights into the data. ARIMA, SARIMA, and LSTM models were generated, evaluated, and visualized to predict NOC 92100 occupational vacancies and analyze macroeconomic relationships with NOC 92100. Study limitations include the

lack of large, available datasets specific to Canadian power engineering, limited existing research literature directly related to power engineering, the choice of exogenous variables as influences on NOC 92100, LSTM limitations with smaller datasets, and potential author bias as a working 1st Class power engineer.

Univariate Data Preprocessing

A single time-dependent variable with a date reference, represented in two individual columns, constitutes a univariate time series. Conversely, multiple time-dependent variables/columns with a date reference constitute a multivariate time series (Ullah et al., 2024). The univariate dataset consists of the date column (REFDATE), arranged at quarterly frequencies, and the NOC 92100 job vacancy column (n = 40), covering the period from 2015 to 2024. The quarterly REFDATE time series was first parsed and indexed as a datetime object to prepare the dataset for exploratory and univariate analysis. The quarterly frequency was confirmed using the resample (QS) function in Python. The univariate NOC 92100 job vacancy data included two missing values attributed to data collection issues during the COVID-19 period. The missing values were imputed using the median value. The median value was calculated using data points immediately surrounding the COVID-19 period to ensure accuracy and continuity. The time-series plot and scatterplot (Figures 1 and 2) show the original quarterly job vacancy trend for NOC 92100 from 2015 to 2024. The histogram shown in Figure 3 indicates a normal distribution for the NOC 92100 occupational vacancies, and the boxplot (Figure 4) and the Q-Q plot (Figure 5) confirmed that no substantial NOC 92100 outliers are present.

Figure 1

NOC 92100 Job Vacancies: Time Series Plot (Original Data)

Figure 2

NOC 92100 Job Vacancies: Scatter Plot (Original Data)

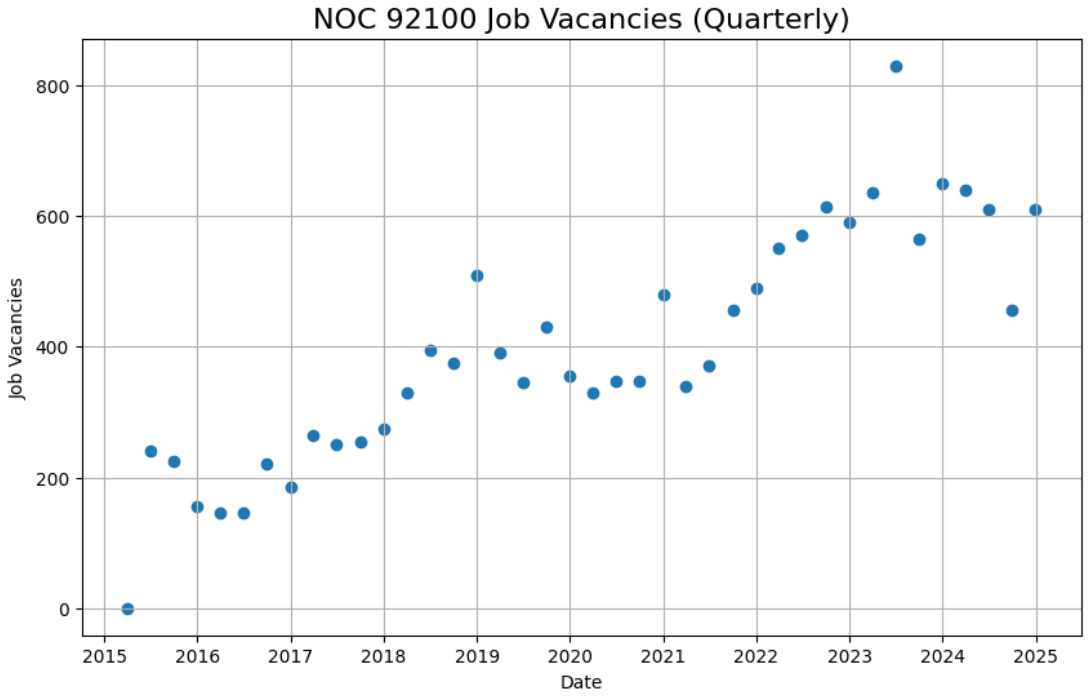


Figure 3

NOC 92100 Job Vacancies: Histogram

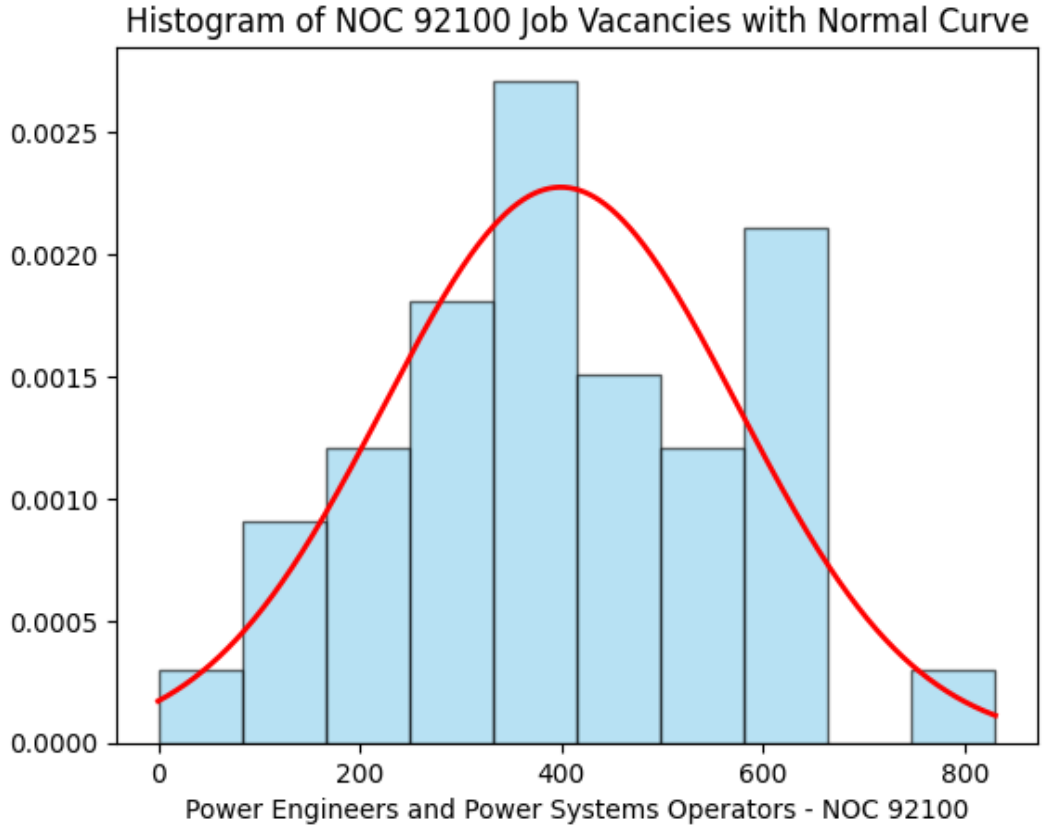


Figure 4
NOC 92100 Job Vacancies: Boxplot

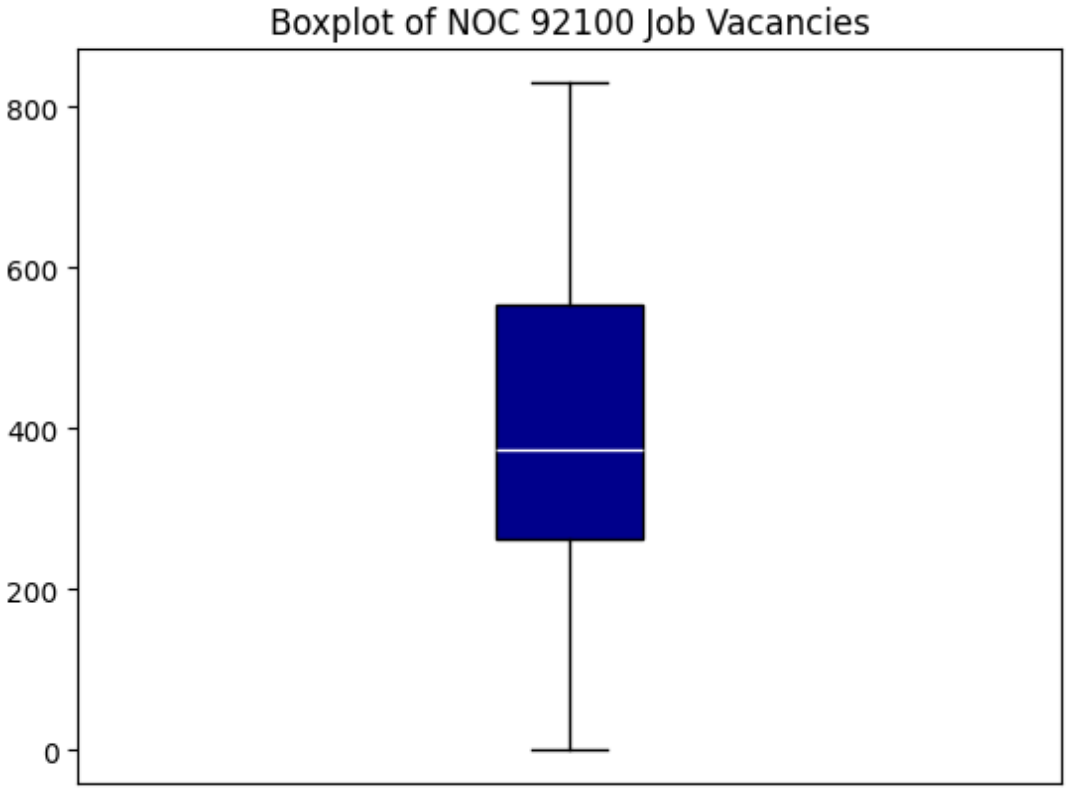
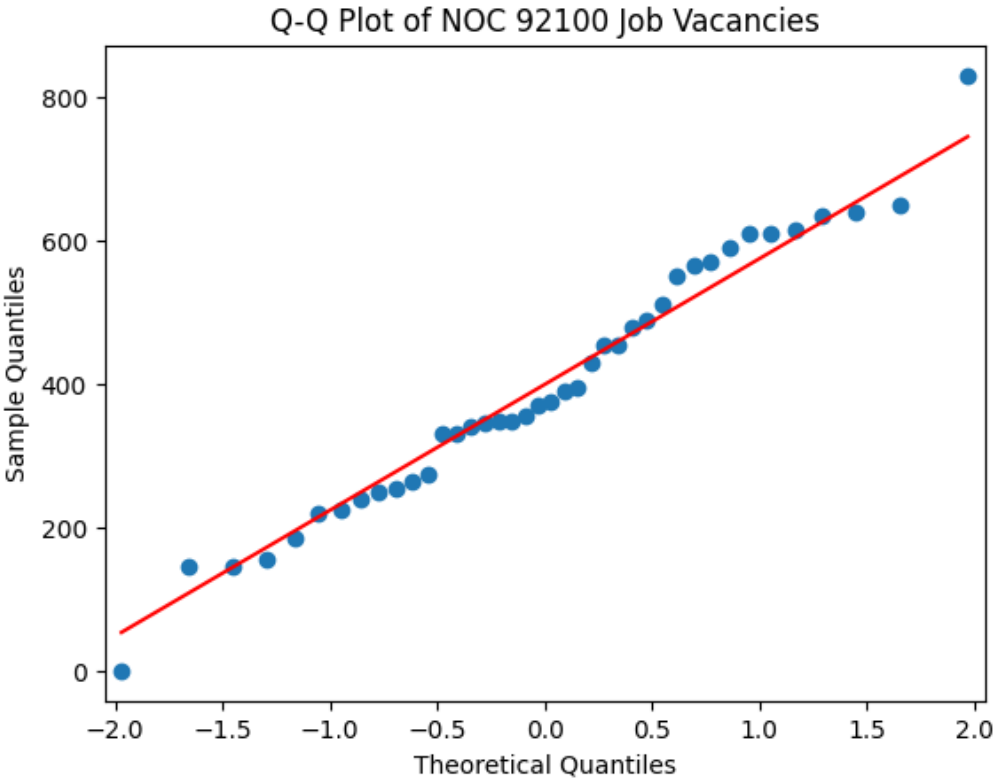


Figure 5
NOC 92100 Job Vacancies: Q-Q Plot

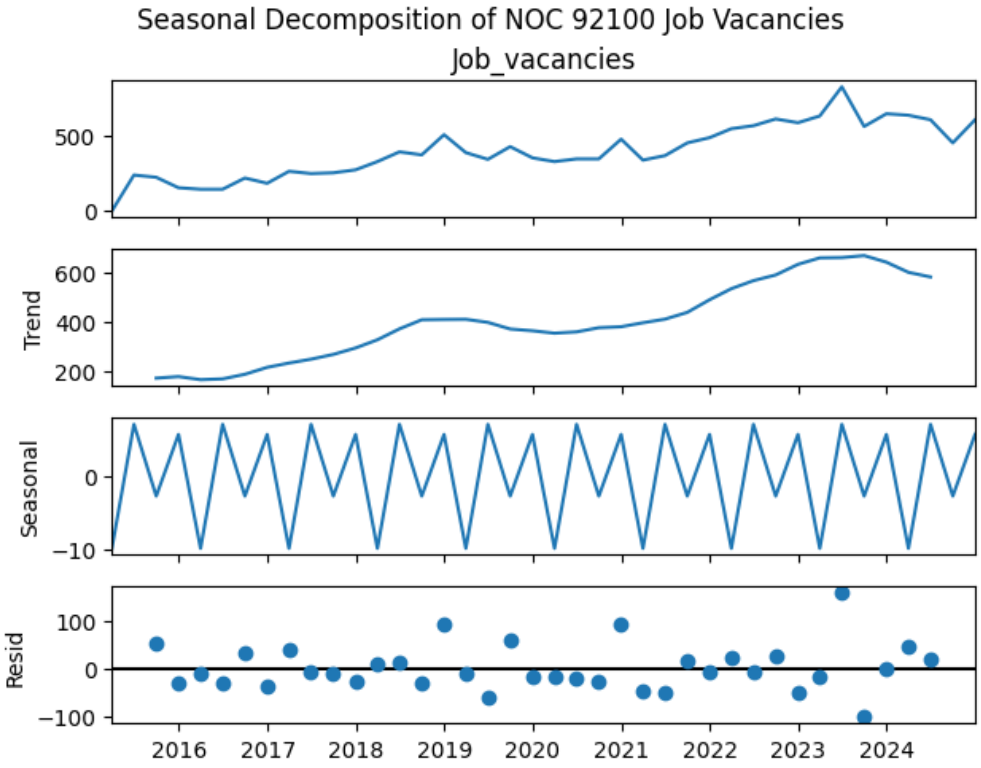


Univariate ARIMA

The NOC 92100 time series was initially checked for stationarity using both statistical and visual methods. A differencing transformation was applied to achieve stationarity and to identify the integration (d) parameter (Atif, 2024; Johnston et al., 2025). It is essential to proactively evaluate and address stationarity and seasonality to ensure compatibility with the chosen time series algorithm and to produce meaningful results (James et al., 2023). Line plots and decomposition graphs offer visual cues regarding the presence or absence of stationarity indicators. Time series decomposition breaks down the data into underlying observations, trend, seasonality, and residual (noise) components (Amalou et al., 2024; Beard et al., 2019). The NOC 92100 job vacancy decomposition plot exhibited a non-stationary trend (Figure 6).

Figure 6

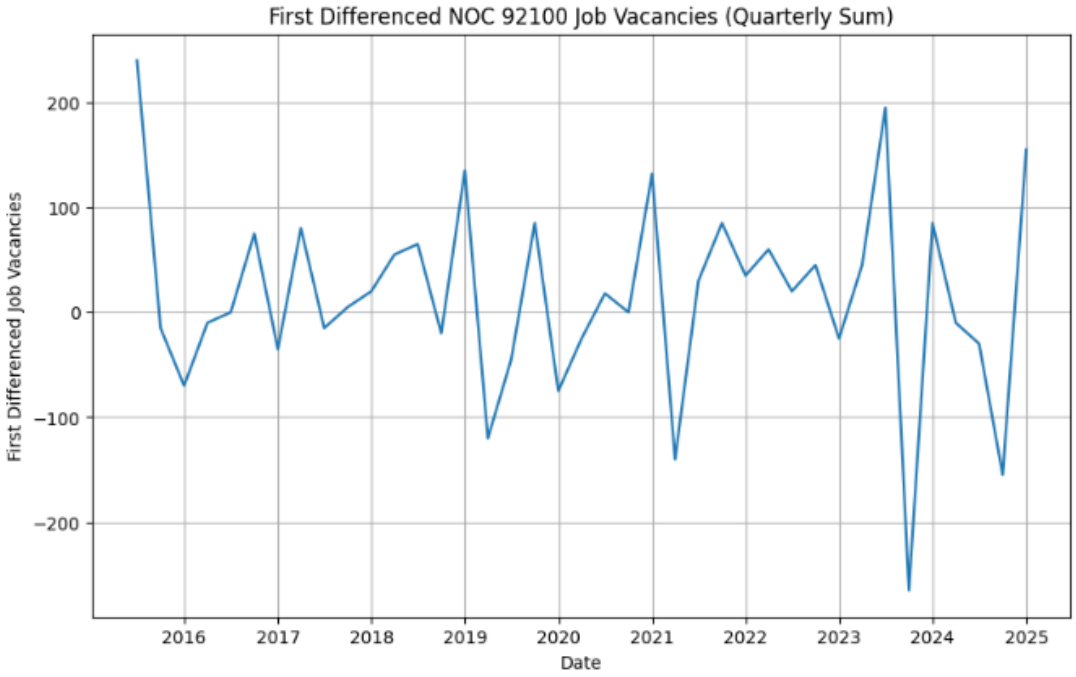
NOC 92100 Job Vacancies: Decomposition Plot



Non-stationarity was further confirmed statistically using the Augmented Dickey-Fuller (ADF) test (Beard et al., 2019; James et al., 2023), which yielded a *p*-value of 0.625, exceeding the 0.05 threshold for non-stationarity. The data were first-differenced to achieve stationarity. This was confirmed by the ADF test, which produced a *p*-value below 0.05. A single differencing operation set the integration parameter (d) to 1. A time series plot was created from the differenced data to confirm stationarity visually (Figure 7).

Figure 7

NOC 92100 Job Vacancies: Time Series Plot (First Differencing)



Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots were produced with the differenced data to visually determine the model parameters 'p' and 'q' (Figures 8 and 9). The ACF measures the correlation between observations separated by different time lags. The PACF measures the correlation between a time series and its lagged values after removing the effects of all intermediate lags. The PACF plot confirmed the 'p' parameter (AR), and the ACF plot confirmed the 'q' parameter (MA) (Cheng et al., 2023; Kontopoulou, 2023). AR and MA were then combined with the differencing result (d) to form the ARIMA model. This process resulted in p, d, q parameters of (1, 1, 1) for the complete ARIMA model. The 'pmdarima' Python package verified the ARIMA parameter selection in a separate notebook. Pmdarima provides automatic fitting of ARIMA and SARIMA models.

Figure 8
NOC 92100 Job Vacancies: ACF Plot

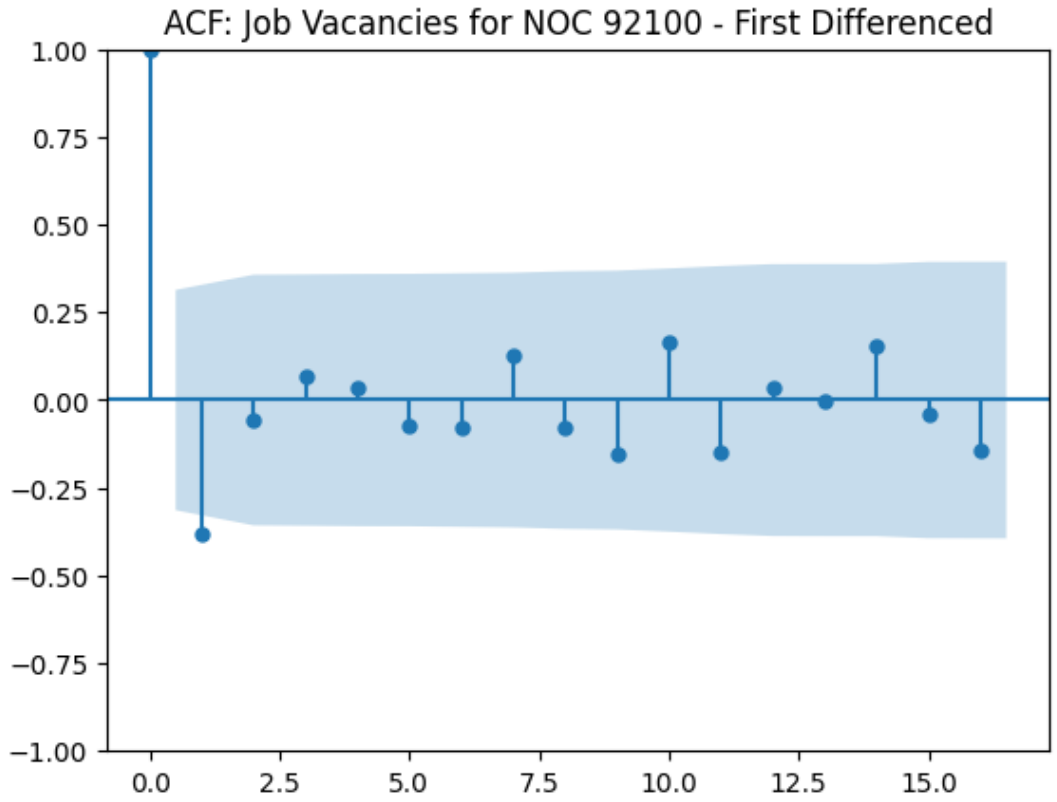
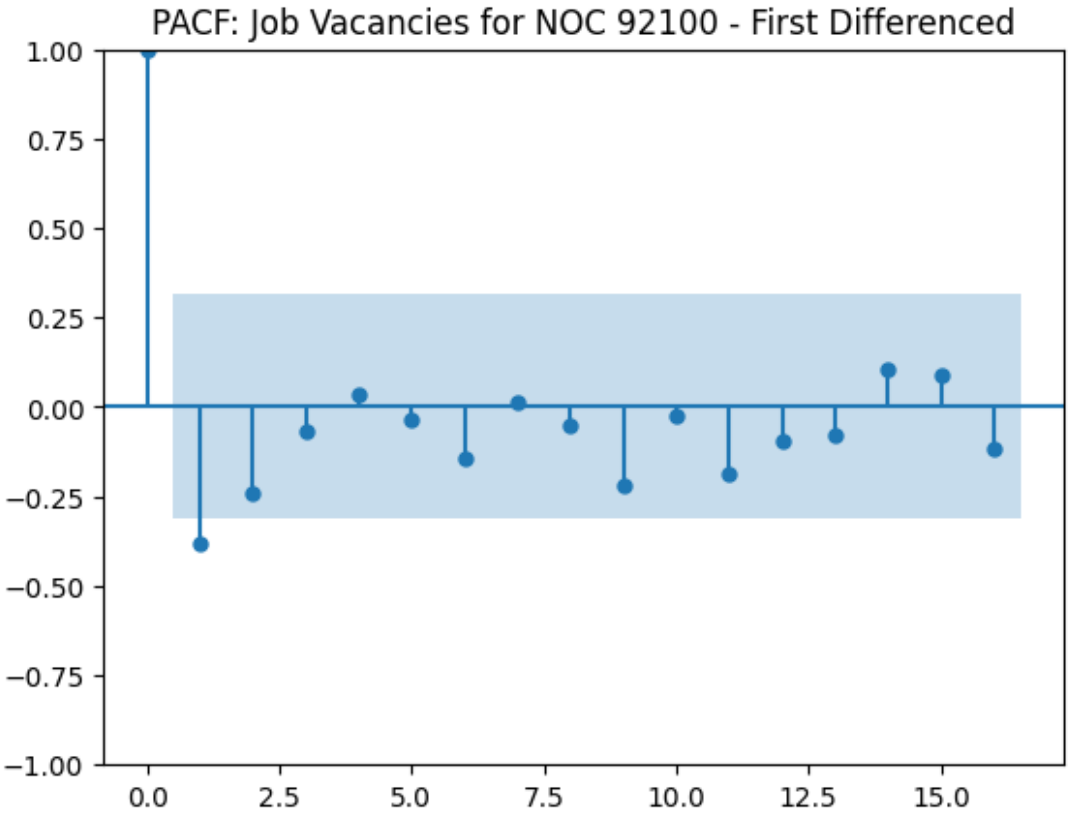


Figure 9

NOC 92100 Job Vacancies: PACF Plot



The model residuals were tested after verifying the model parameters to assess whether the model aligns with white noise (Cheng et al., 2023). Residuals were

calculated as the difference between the observed actual values and the predicted values from the fitted model. Residual diagnostics evaluate the model's fit and identify potential issues, such as bias or patterns, that could affect the forecast. Based on the histogram and Q-Q plot (Figures 10 and 11), the residuals appear approximately normally distributed, with most clustering around zero. The Q-Q plot indicates that the residuals generally follow a straight line, suggesting normality, although some deviations occur at the tails.

Figure 10

NOC 92100 Job Vacancies: Histogram Residuals on Fitted Model

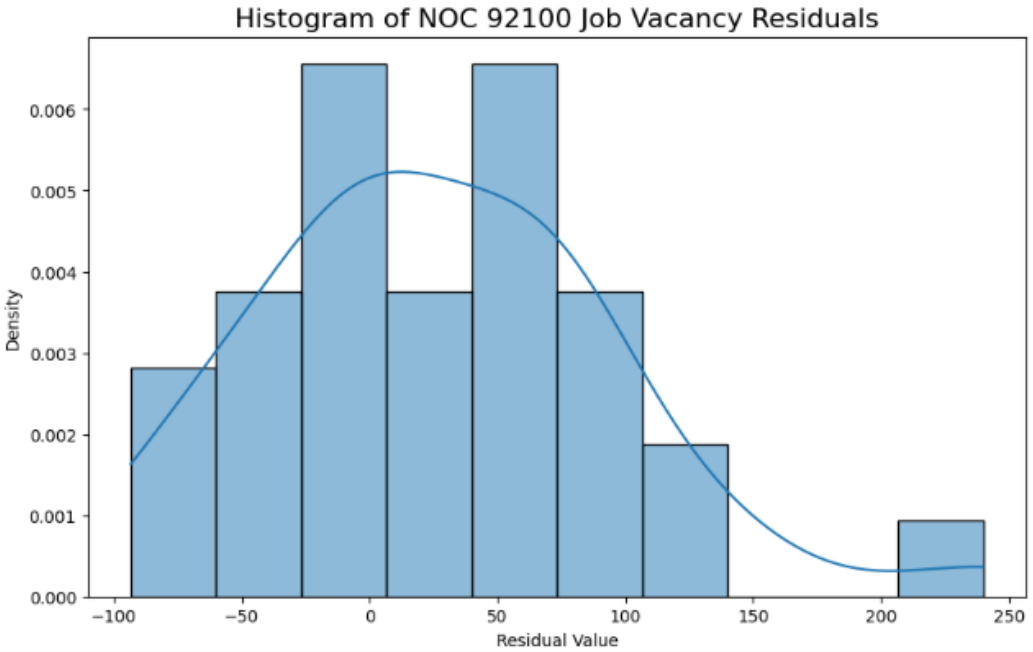
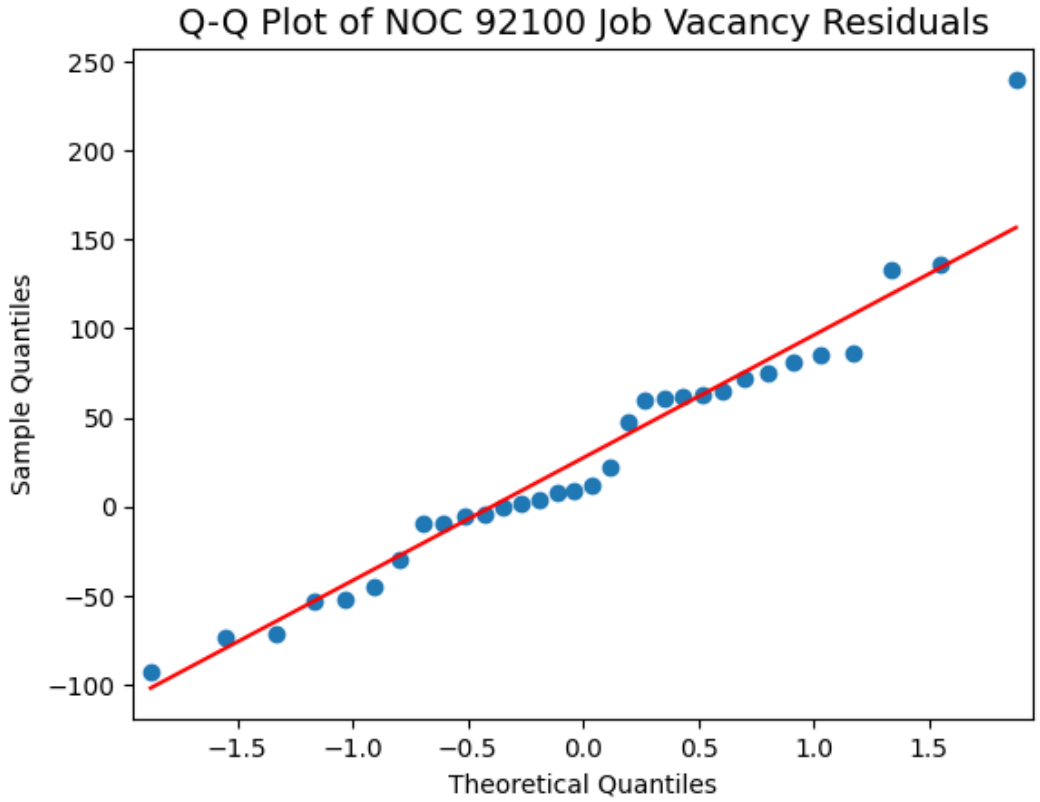


Figure 11

NOC 92100 Job Vacancies: Q-Q Plot Residuals on Fitted Model



The ARIMA time series data was split into 80% for training and 20% for testing. The ARIMA/SARIMA sections use an 80/20 split and a fixed test horizon, such as eight steps. The ARIMA model was built using the training data, with predictions made on the test data. The model's accuracy was then evaluated using RMSE, MAE, and SMAPE.

Univariate SARIMA

The univariate NOC 92100 dataset used for the ARIMA model was also utilized for the SARIMA model, employing the same preprocessed data. SARIMA is a type of regression analysis that incorporates seasonality into autoregressive moving-average models. SARIMA model estimates use maximum likelihood estimation (MLE) to identify parameters that maximize the likelihood of the observed data (Gopu et al., 2023). The SARIMA model extends the ARIMA model by managing seasonality in univariate time series data. SARIMA combines the non-seasonal ARIMA p,d,q orders with seasonal P,D,Q,s orders. The seasonal order includes AR (P), differences (D), MA (Q), and periodicity (s) terms (Sirisha et al., 2022). The number of periods per season is represented by (s). This process resulted in p, d, q parameters of (1, 1, 1) for the overall ARIMA model. The order of the SARIMA model is generally not high (Selmy et al., 2024). The 'pmdarima' Python package verified the SARIMA (P, D, Q) parameter selection as 0, 0, 0 with quarterly periodicity (4). Pmdarima was used in a separate notebook to confirm the SARIMA model orders, which are consistent with the ARIMA model. The full SARIMA hyperparameters are (1, 1, 1) (0, 0, 0, 4). The P,D,Q SARIMA orders (0, 0, 0) produce the exact plotted prediction and evaluation results as the ARIMA model output.

Univariate and Multivariate LSTM Prediction Models

Univariate and multivariate predictive LSTM models were developed in this study for comparison with each other and with the stochastic ARIMA and SARIMA models. Only the multivariate LSTM model also evaluated the relationship between exogenous macroeconomic factors and NOC 92100 vacancies. The univariate LSTM predictive model

used the same NOC 92100 preprocessed dataset as the ARIMA and SARIMA models, comprising job vacancies (n = 40) with quarterly frequencies spanning 2015 to 2024. The multivariate LSTM predictive model utilized the expanded univariate dataset, comprising 122 NOC 92100 observations (rows) and 30 measures (columns), covering the years 2015 to 2024. The dataset comprises the NOC 92100 occupational vacancy counts, and related Statistics Canada measures derived from labor force characteristics, employment by industry, interest rates, consumer price index (CPI), inflation, household income, GDP income, and GDP NAICS. The NOC 92100 multivariate dataset was first checked for outliers using the +/- standard deviation (SD) and interquartile range (IQR) methods. These tests indicated that no significant outliers were present. The missing values in the multivariate dataset were evaluated and imputed using the *iterative.imputer* function in the Python Scikit-learn (sklearn) package (Domino, 2025). The univariate and multivariate models required optimization after their initial runs. Optimization was first carried out on the univariate model. Optimization insights obtained from the univariate trials were then applied to the multivariate model. The best performance, as measured by RMSE, MAE, and SMAPE, was achieved at the 11th optimization cycle, after which performance decreased. The model was constructed using 'Keras' and the Sequential, Dense, and LSTM functions to configure the neural network. Keras is the high-level API in TensorFlow that streamlines the generation, training, and evaluation of machine learning models (Keras, 2025). TensorFlow is a flexible and powerful library for building and training models, primarily when used with Python and Keras (Keras, 2025; TensorFlow, 2025). The models were constructed with the following parameters:

- LSTM Units = 50 (output space dimensionality for both univariate and multivariate models)
- Lookback Window = 20 (number of sequential historical data points to make a prediction)
- Dropout layers = The standard dropout is 0.2, and the recurrent dropout = 0.0
- Dense layer = 1 (fully connected NN layer including L2 regularization 1e-4)
- Optimizer = Adaptive Moment Estimation (Adam) with learning rate (3e-4) and Huber loss (delta=0.05)
- Epochs = Number of epochs for the model training (univariate = 50 and multivariate = 50)
- Batch size = Number of samples per gradient update (univariate = 16 and multivariate = 16)
- Early stopping = Univariate (patience-based) and multivariate (threshold-based)
- Verbose = 2 (controls the level of output displayed during the training process)
- Future steps = The univariate model extends to a calendar end date. These future steps are calculated from the last observed quarter to December 31, 2027. The multivariate model uses eight (8) future steps.

A series of methodological adjustments was implemented to optimize the training of recurrent neural networks (RNNs) for quarterly time-series forecasting. First, data preprocessing was improved by fitting the MinMaxScaler exclusively to the training set to prevent data leakage. Temporal context was expanded by extending the lookback window from 16 (four years) to 20 (five years) quarterly observations ('QS' in the forecast code), thereby enabling the model to capture slower cyclical dynamics and regime shifts (Zhu et al., 2025). Epochs were set at 50. Initially, epoch iterations were terminated when a predefined threshold was reached. This approach was later revised to terminate training at the epoch with the minimum validation loss among the 50 total epochs. Consequently, iterations only reached the final (50th) epoch when, under the original threshold-based method, the minimum validation loss had not yet been achieved. Under the revised minimum validation loss criterion, all 50 epochs were completed only if the lowest

validation loss occurred at the final epoch. This reduces overfitting (Bi & Neethirajan, 2024) and improves MAE and SMAPE. To better preserve sequential dependencies, training was conducted without shuffling. This prevented the mixing of early- and late-regime windows within epochs. The Huber loss function delta was reduced from 1.0 to 0.05. This adjustment balanced the quadratic sensitivity for minor errors with a linear treatment of larger residuals, thereby improving robustness to outliers (Sun et al., 2025). A light layer 2 (L2) weight regularization was applied to the output layer to reduce variance and stabilize predictions. The kernel regularizer was included in the final dense layer. The batch size was reduced from 32 to 16, introducing beneficial gradient noise that acted as a regularizer in time-ordered sequences (Mishra et al., 2024). Optimization refinements included lowering the Adaptive Moment Estimation (Adam) learning rate to minimize oscillations around the validation minimum and initially introducing recurrent dropout (0.10) to regularize temporal dependencies. However, subsequent trials revealed that recurrent dropout was overly restrictive for this dataset. Removing the recurrent dropout (0.0) restored efficient cuDNNLSTM execution, strengthened and attenuated medium-term temporal learning, and improved RMSE performance on noisy small-sample series (Sangeetha et al., 2024). Overall, these adjustments collectively reduced MAE, SMAPE, and RMSE by enhancing generalization, stabilizing the optimization process, and tailoring regularization to the idiosyncratic dynamics of macroeconomic vacancy series.

Results

Univariate ARIMA Prediction Model

An ARIMA prediction model does not account for the negligible seasonality in the time series data. The prediction indicates a flat trend for NOC 92100 occupational vacancies at 600 (Figures 12 and 13) with RMSE, MAE, and SMAPE at 102.43, 72.99, and 11.74%, respectively (Table 1).

Figure 12

NOC 92100 Job Vacancies: ARIMA Time Series Plot

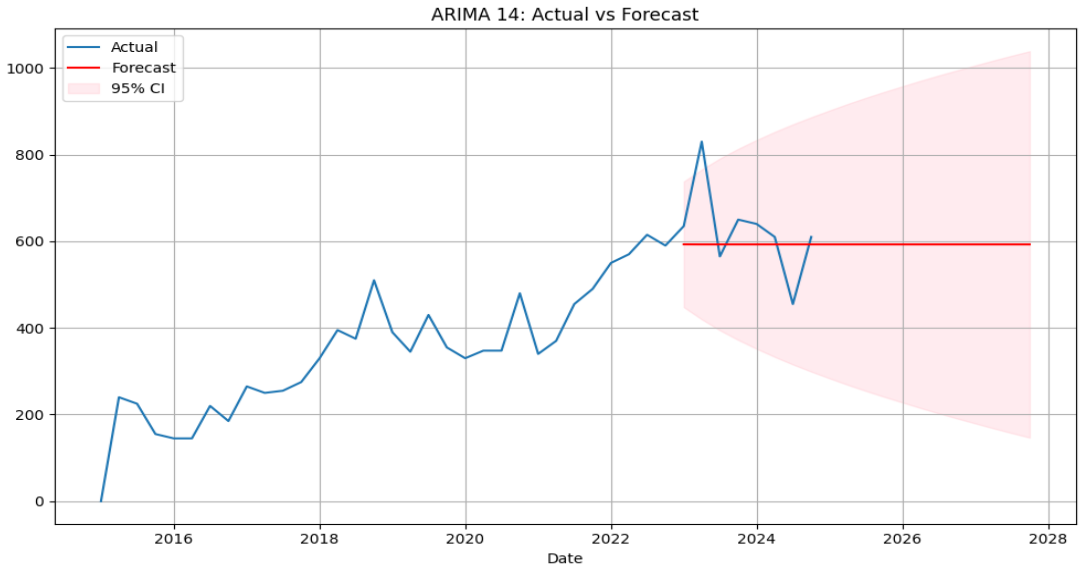


Figure 13

NOC 92100 Job Vacancies: ARIMA Time Series Sarimax Output

SARIMAX Results

Dep. Variable:	outcome	No. Observations:	32
Model:	ARIMA (1,1,1)	Log Likelihood	-177.495
Date:	Sun, 24 Aug 2025	AIC	360.990
Time:	16:28:45	BIC	363.292
Sample:	01-01-2015	HQIC	362.392
	10-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0600	0.674	-0.089	0.929	-1.381	1.261
ma.L1	-0.2924	0.624	-0.469	0.639	-1.515	0.930
Sigma2	5481.4476	1322.790	4.144	0.000	2888.826	8074.069

Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	1.94
Prob(Q):	0.91	Prob(JB):	0.38
Heteroskedasticity (H):	0.72	Skew:	0.56
Prob(H) (two-sided):	0.61	Kurtosis:	3.51

Univariate SARIMA Prediction Model

The SARIMA model produced a similar prediction to the ARIMA model, with seasonal orders of 0, 0, 0, 4. The prediction again indicates a flat trend for NOC 92100 occupational vacancies at 600 (Figures 14 and 15) with RMSE, MAE, and SMAPE at 102.43, 72.99, and 11.74%, respectively (Table 1).

Figure 14

NOC 92100 Job Vacancies: SARIMA Time Series Plot

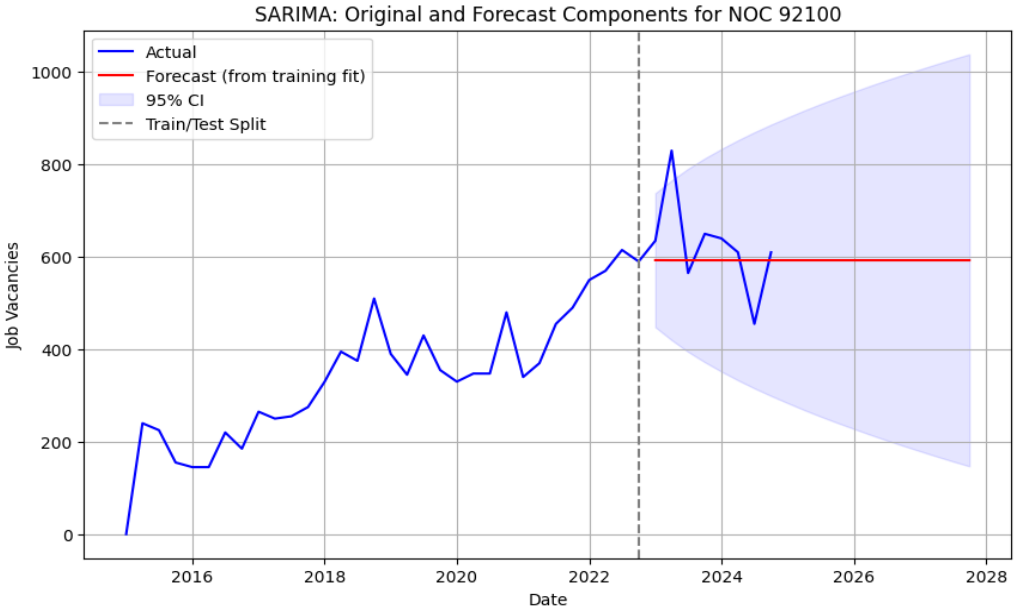


Figure 15

NOC 92100 Job Vacancies: SARIMA Time Series Sarimax Output

SARIMAX Results

Dep. Variable:	outcome	No. Observations:	32
Model:	SARIMAX (1,1,1)	Log Likelihood	-177.495
Date:	Sun, 24 Aug 2025	AIC	360.990
Time:	16:49:29	BIC	363.292
Sample:	01-01-2015	HQIC	362.392
	10-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	-0.0600	0.674	-0.089	0.929	-1.381	1.261
ma.L1	-0.2924	0.624	-0.469	0.639	-1.515	0.930
Sigma2	5481.4476	1322.790	4.144	0.000	2888.826	8074.069

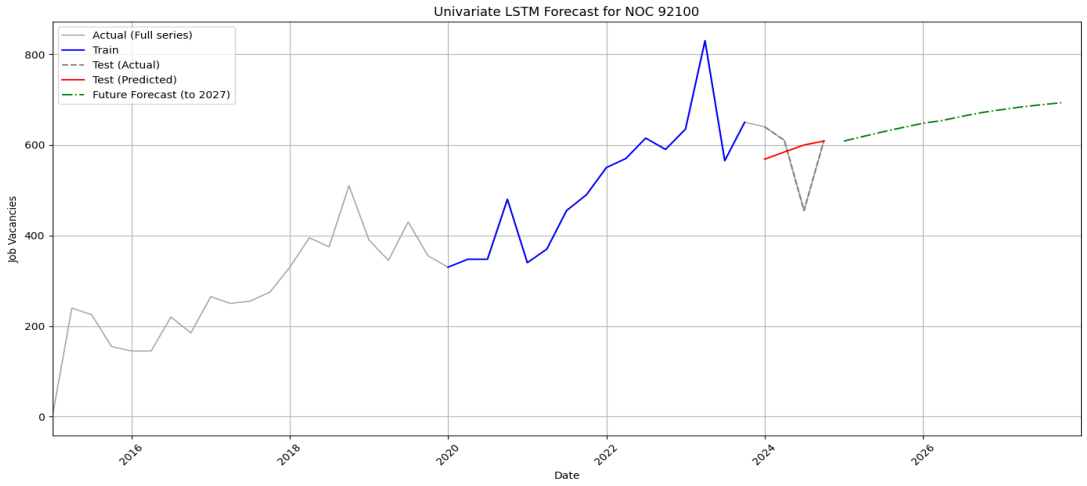
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	1.94
Prob(Q):	0.91	Prob(JB):	0.38
Heteroskedasticity (H):	0.72	Skew:	0.56
Prob(H) (two-sided):	0.61	Kurtosis:	3.51

Univariate LSTM Prediction Model

The univariate prediction indicates NOC 92100 job vacancies trending upwards from 600 to 700 (Figure 16) with RMSE, MAE, and SMAPE at 82.16, 59.51, and 10.72%, respectively (Table 1).

Figure 16

NOC 92100 Job Vacancies: LSTM Univariate Time Series Plot

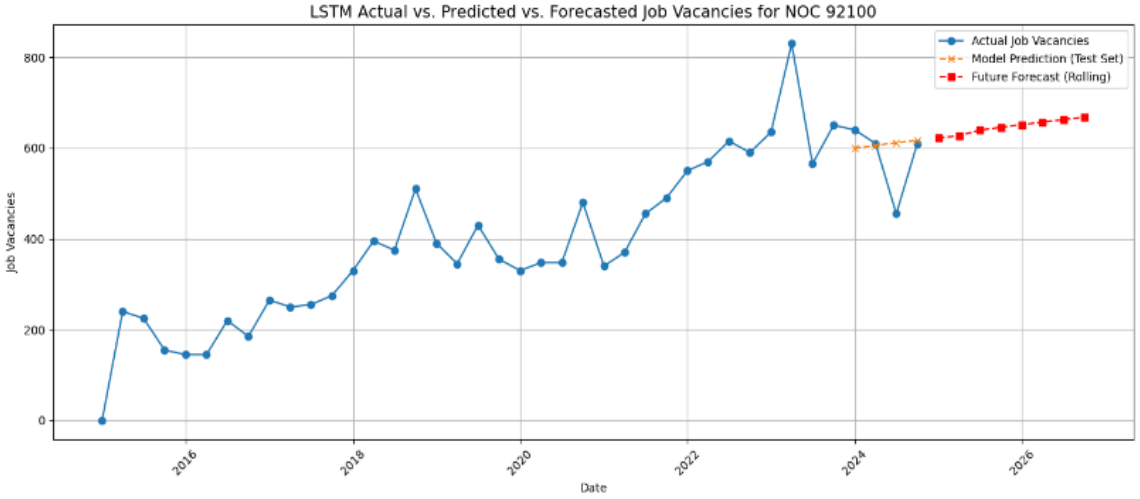


Multivariate LSTM Prediction Model

The multivariate prediction indicates NOC 92100 job vacancies trending upward from 600 to 700 (Figure 17) with RMSE, MAE, and SMAPE at 79.02, 50.45, and 9.17%, respectively (Table 1). Multivariate analysis differs from univariate analysis by including additional macroeconomic exogenous measures.

Figure 17

NOC 92100 Job Vacancies and Predictors: LSTM Multivariate Time Series Plot



Model Evaluation Metrics

Evaluation metrics are critical for making relative comparisons and drawing conclusions among predictive models. RMSE, MAE, and MAPE are three reliable metrics for evaluating models. RMSE indicates the average difference between the actual values in the dataset and the model's predictions. This metric essentially compares the dataset's estimated and actual values. The model's ability to fit the data is inversely proportional to the RMSE value. MAE is the cumulative average of the absolute errors, representing how a model replicates a full dataset. The metric aggregates the deviations between the predicted and expected values as the numerator and divides this sum by the total number of predictions. MAPE is the average absolute percentage error for each prediction minus the original values as the numerator divided by the actual values (Gopu et al., 2023; Selmy et al., 2024). SMAPE is a version of MAPE that uses the average of the actual and forecasted values in its denominator. SMAPE ranges from 0% to 100%, with lower values signifying higher accuracy (Kolassa, 2025).

Table 1

Model Evaluation Metrics

	RMSE	MAE	SMAPE	Parameters
ARIMA Univariate	102.43	72.99	11.74%	(1,1,1)
SARIMA Univariate	102.43	72.99	11.74%	(1,1,1) (0,0,0,4)
LSTM Univariate	82.16	59.51	10.72%	LSTM Units = 50 Lookback window = 20 Dropout layer = 0.2 Recurrent = 0.0 Dense layer = 1 Optimizer = Adam Epochs = 50 maximum Batch size = 16 Early stopping = patience-based Verbose = 2

LSTM Multivariate	79.02	50.45	9.17%	Future steps = December 31, 2025 LSTM Units = 50 Lookback window = 20 Dropout layer = 0.2 Recurrent = 0.0 Dense layer = 1 Optimizer = Adam Epochs = 50 Batch size = 16 Early stopping = patience-based Verbose = 2 Future steps = 8

Multivariate LSTM Exploratory Factor Analysis and Correlations

The multivariate sample added 28 variables (measures) to the quarterly NOC 92100 univariate vacancy dataset as potential predictors of NOC 92100 vacancies. The 28 measures were derived from Statistics Canada's categories of labor force characteristics, employment by industry, interest rates, consumer price index (CPI), inflation, economic household income, GDP income, and GDP NAICS. The broad range of measures was chosen to gain a general understanding of the factors influencing NOC 92100 vacancies in this study. The multivariate time-series analysis was followed by an exploratory factor analysis (EFA) of the 28 variables to identify the latent factors most strongly influencing NOC 92100 occupational vacancies. EFA was conducted with the Python *FactorAnalyzer* module and comprised Varimax rotation, factor loadings, a Scree plot of eigenvalues with an explained variance ratio, Pearson's *r*, and Spearman's *rho* correlational analysis, and Steiger's Z test to evaluate the statistical significance of the differences between dependent correlations. Factor loadings are displayed in Table 2.

Table 2

Exploratory Factor Analysis – Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
SS Loading	12.186	11.370	2.482	1.069	0.426	0.424
Proportional Var	0.435	0.406	0.089	0.038	0.015	0.001
Cumulative Var	0.435	0.841	0.930	0.968	0.983	0.984

Note. The hypothesis test indicated that six factors are sufficient, with four factors retained.

The chi-square statistic is 2128.87 on 225 degrees of freedom.

The *p*-value is < 0.001

The four predictors include a wealth index, a real household economic activity index, a core inflation composite index, and a labor market slack indicator, after dimension reduction (Table 3). Pearson's *r* and Spearman's *rho* correlations are displayed in Table 3.

Table 3
Exploratory Factor Analysis – Dimension Reduction

Measure	Factor Loading			
	1	2	3	4
Average offered hourly wage	0.191	0.767	0.232	0.098
Assets as a share of net worth	**0.945**	0.297	0.131	0.010
Debt to disposable income ratio	**0.944**	0.290	0.148	0.016
Financial assets as a share of net worth	**0.956**	0.274	0.091	0.017
Financial assets as a share of total assets	**0.953**	0.283	0.094	0.009
Financial assets to nonfinancial assets ratio	**0.961**	0.259	0.053	0.022
Liabilities as a share of net worth	**0.957**	0.246	0.115	0.045
Liabilities as a share of total assets	**0.955**	0.257	0.118	0.038
Net worth as a share of disposable income	**0.920**	0.340	0.166	0.026
Nonfinancial assets as a share of total assets	**0.928**	0.327	0.175	0.004
Owners' equity as a share of real estate	**0.936**	0.318	0.148	0.011
Real estate as a share of disposable income	**0.901**	0.358	0.209	0.027
Real estate as a share of total assets	**0.923**	0.335	0.181	0.011
Employment by industry count	0.381	**0.900**	0.189	0.047
Total funds advanced for residential mortgages insured	0.029	**0.741**	0.118	0.279
Core inflation factor CPI common price changes	0.264	0.461	**0.806**	0.182
Core inflation weighted median CPI median - 50th percentile	0.304	0.582	**0.732**	0.144
Core inflation trimmed mean CPI - no extremes	0.302	0.517	**0.771**	0.150
Core inflation weighted median CPI index	0.361	**0.916**	0.157	0.004
Core inflation trimmed median CPI - no volatility	0.357	**0.914**	0.177	0.012
CPI excluding the eight most volatile components	0.355	**0.906**	0.207	0.034
CPI excluding the eight most volatile components - SA	0.350	**0.908**	0.210	0.035
GDP by income - compensation of employees	0.347	**0.916**	0.157	0.106
GDP by income - wages and salaries	0.347	**0.916**	0.159	0.106
GDP at annual rates using chained 2017 dollars - SA	0.374	**0.880**	0.254	0.125
GDP at annual rates using 2017 constant prices - SA	0.375	**0.881**	0.254	0.119
GDP trading day adjusted using 2017 constant prices	0.364	**0.886**	0.252	0.109
Unemployment related to the NOC	0.069	0.257	-0.231	**0.931**

SA = seasonally adjusted.

The extraction method was an exploratory factor analysis with Varimax rotation. Factor loadings above .50 (absolute value) are displayed in bold.
Factor loading 1 = Wealth Index (WI); Factor loading 2 = Real Household Economic Activity (RHEAI) Index; Factor loading 3 = Core Inflation Composite Index (CICI); Factor loading 4 = Labour Market Slack Indicator (LMSI)
EFA dimension reduction reduced the 28 measures to four aggregated factors, as indicated by the Scree plot's 'elbow', where the slope decreases significantly (Figure 18). The four factors displayed reasonable normality and linearity.

Figure 18

Exploratory Factor Analysis – Scree Plot

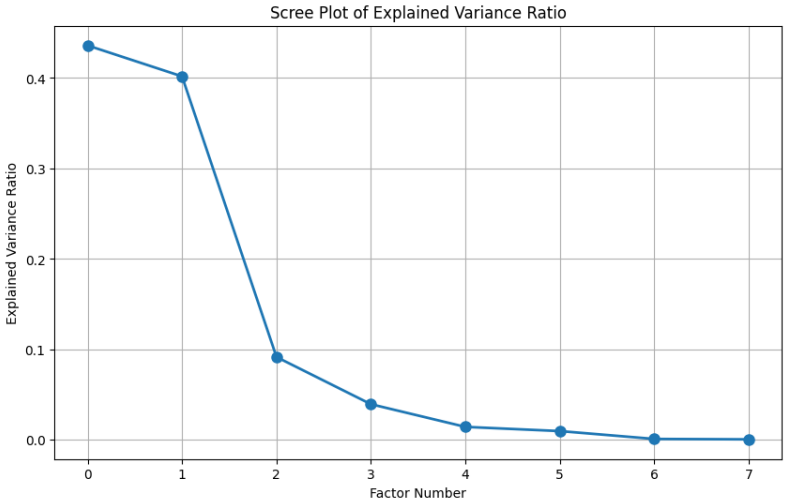


Table 4 indicates that the Pearson's *r* correlations for Factor 1 (WI) and Factor 2 (RHEAI) are 0.327 (*p*-value < 0.05) and 0.776 (*p*-value < 0.05), respectively. Factor 2 (RHEAI) was the only significant Spearman's *rho* correlation with a coefficient of 0.849 (*p*-value < 0.05). Factor 3 (CICI) and Factor 4 (LSMI) correlations are not significant for either Pearson's *r* or Spearman's *rho* (Table 4). However, Pearson's *r* correlation for Factor 3 (CICI) approached significance at 0.288 (*p*-value = 0.07).

Table 4

Exploratory Factor Analysis – Factor Correlations with Job Vacancies (Dependent Variable)

	Predictor	Pearson's *r*	Pearson's *p*-value	Spearman's *rho*	Spearman's *p*-value
0	Factor 1	0.327	0.0397	0.206	0.2015
1	Factor 2	**0.776**	**0.0000**	**0.849**	**0.0000**
2	Factor 3	0.288	0.0719	0.097	0.5530
3	Factor 4	0.103	0.5256	0.125	0.4416

Steiger's Z evaluated the statistical significance of differences in dependent correlations to determine whether the strength of the correlation between two factors is significant (Table 5). A consideration of the test is the smaller sample size (n = 40).

Table 5

Steiger's Z Test – Correlations between Factors

Factors Compared	Steiger's Z	*p*-Value	Significance
F1 vs **F2**	-1.935	0.0530	**Approaching**
F1 vs F3	0.167	0.8672	NO
F1 vs F4	0.960	0.3371	NO
F2 vs F3	2.102	0.0356	**YES**
F2 vs F4	2.895	0.0038	**YES**
F3 vs F4	0.793	0.4279	NO

Factors 1 vs 2: Correlations of 0.327 and 0.776 with job vacancies (DV), respectively.
Factors 2 vs 3: Correlations of 0.776 and 0.288 with job vacancies (DV), respectively.
Factors 2 vs 4: Correlations of 0.776 and 0.103 with job vacancies (DV), respectively.
All four factors are retained, with Factor 2 having a significant influence on job vacancies (DV). Feature ablation was used as a supplementary method alongside Pearson's *r* to evaluate feature importance. Pearson's *r* runs one variable at a time while feature ablation first runs the model with four factors, then reruns the model with three factors,

and measures the change in RMSE due to the removal of the specified factor (Table 6 and Figure 9). Feature ablation relies on the non-deterministic LSTM analysis (Li & Li, 2025).

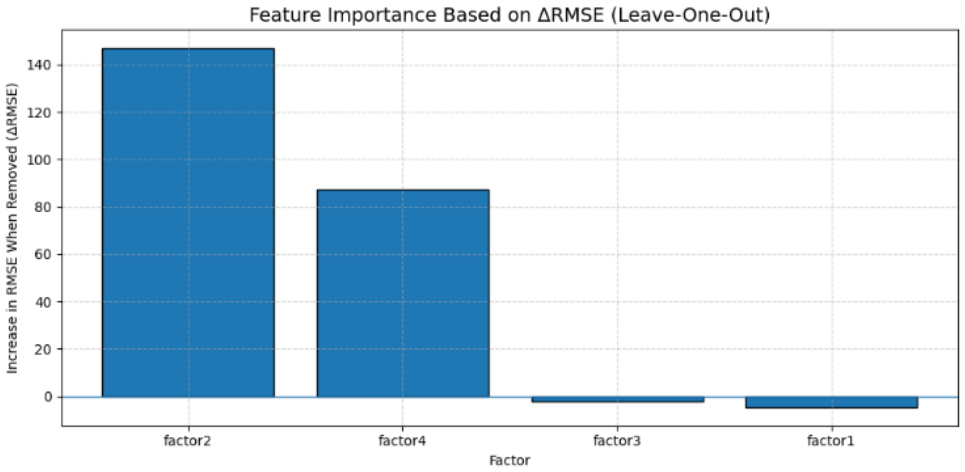
Table 6

Feature Ablation – Feature Importance Ranking

Factors	RMSE	Base RMSE (all features)	Δ RMSE
Factor 2	228.08	79.02	149.06
Factor 4	168.18	79.02	89.16
Factor 3	78.74	79.02	-0.28
Factor 1	76.09	79.02	-2.93

Figure 9

Feature Ablation – Feature Importance Ranking



Discussion

This quantitative, exploratory study provides estimates of future NOC 92100 occupational vacancies and the key factors influencing them (predictors). Its aim was to identify, conservatively, the most effective methods for forecasting vacancies and the factors that affect them. The specific numerical results from the time series models and relationship analysis are less important than the insights gained into choosing the most appropriate methodologies. These high-level insights will support future research and may help job seekers make better-informed career decisions. The original NOC 92100 vacancy data (2015–2024) indicate that vacancies increased compared to previous periods, with a notable peak of 830 vacancies during the COVID-19 pandemic (Gil-Alana et al., 2025; Mavrodi et al., 2024) and then stabilized during the prediction period (2025–2027) (Figure 1). ARIMA, SARIMA, and LSTM time series forecasting models predict a future NOC 92100 vacancy range of 600–700 from 2025 to 2027. Pearson's *r* (r = 0.776, $p < 0.05$) and Spearman's *rho* ($r = 0.849$, $p < 0.05$) analysis of the relationships between the exogenous macroeconomic measures (predictors) and NOC 92100 vacancies (DV) revealed a strong and significant correlation between Factor 2 (RHEAI) and NOC 92100 vacancies. Feature ablation supported these findings, with Factor 2 being the most influential, resulting in a 149.06-point change in RMSE (Table 6). Factor 2 represents Real Household Economic Activity (RHEAI) and includes strong loadings on residential mortgage loans, core inflation and CPI indexes, wages and salaries, and GDP levels (Table 3). The multivariate LSTM model outperformed the univariate LSTM model and the ARIMA and SARIMA statistical models across RMSE, MAE, and SMAPE. A review of the literature confirmed the robustness and adaptability of LSTM-based models across various domains and data structures, particularly in multivariate and long-term time series prediction. While traditional models like ARIMA and SARIMA retain relevance

in certain applications, the superior performance of LSTM in handling nonlinearity, long-range dependencies, and complex variable interactions positions LSTM as a preferred method for high-accuracy forecasting in both academic and applied research contexts (Gopu et al., 2023; Selmy et al., 2024; Sirisha et al., 2022; Ullah et al., 2024).

Conclusion

Using the multivariate LSTM with an eight-quarter forecast horizon (future steps = 8), the held-out evaluation yields RMSE = 79.02, MAE = 50.45, and SMAPE = 9.17% (Table 1), and the forecast indicates NOC 92100 job vacancies increase from approximately 600 to 700 over the 2025–2027 prediction window (Figure 17). Across all forecasted quarters, the 95% prediction intervals remain above the recent baseline, indicating a persistent upward trend. Feature ablation shows Factor 2 (Real Household Economic Activity Index) is the dominant driver, with removal increasing error by ΔRMSE = +149.06 (Table 6). Factor 2 (RHEAI) is most important (ΔRMSE +149.06), followed by Factor 4 (ΔRMSE +89.16). Removing Factor 2 or Factor 4 degrades performance, while removing Factor 1 or Factor 3 has a negligible effect (and may slightly improve RMSE). In rolling backtests, the LSTM outperforms ARIMA and SARIMA. From a workforce and industry perspective, power engineers are well-placed to transfer their extensive skills to other sectors. Increased visibility and advocacy are likely to influence talent inflows and help clarify interdisciplinary differences (Ahn et al., 2023). Expanding outreach and defining modern role profiles can improve career pathways and enhance diversity among entrants. Power engineers are an underutilized resource for industrial and economic growth, and recruitment can foster cross-sector and cross-functional integration (Huang et al., 2022; Pimenta et al., 2016). Power engineering outcomes could only be inferred from NOC 92100 data, which also includes power systems operators. Study limitations include structural issues related to power engineering certification levels and wage premiums. Power engineering vacancies are tied to certification levels, so vacancies may remain if a candidate with the required certification is unavailable or if the advertised wage is unappealing. Limitations also arise from using exogenous series as predictors, as well as from issues such as concept drift, data non-stationarity, and the chosen prediction horizon. The identified factor relationships are correlational, not causal. Scaling effects across units, such as dollars versus thousands of dollars, impact the model coefficients. Future research could explore alternative exogenous indicators to predict occupational vacancies and investigate whether vacancies stem from shortages of qualified engineers, credentialing bottlenecks, or cross-sector migration driven by employability perceptions and social comparisons, using job posting, application, and wage data. Other machine learning methods using rolling backtests with confidence intervals for ΔSMAPE/ΔRMSE/ΔMAE should be examined. Model-agnostic methods may improve the interpretability of results across different models. Exploring correlations between sectoral total factor productivity (Solow residual) and vacancy dynamics, rather than causal inference, will differentiate economic growth from automation beyond increases in labor and capital inputs. Comparing vacancy forecasts for NOC 92100 with those of other engineering-related disciplines presents another promising avenue. Finally, benefits will accrue from demystifying the myopic *naïve view of power engineering*.

References

ABSA. (2026). *The pressure equipment safety authority*. https://www.absa.ca/

Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Artificial intelligence: The ambiguous labor market impact of automating prediction. *Journal of Economic Perspectives*, *33*(2), 31–50. https://doi.org/10.1257/jep.33.2.31

Ahn, Y., Yan, M., Lin, Y.-R., Chung, W.-T., & Hwa, R. (2023). Tribe or not? Critical inspection of group differences using TribalGram. *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. https://doi.org/10.1145/3484509

Alharbi, F., & Al-Alawi, A. I. (2024, January). Labor market prediction using machine learning methods: A systematic literature review. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 478–482). IEEE. https://doi.org/10.1109/ICETSIS61505.2024.10459632

Amalou, I., Mouhni, N., & Abdali, A. (2024, April). CNN-LSTM architectures for non-stationary time series: Decomposition approach. In *2024 International Conference on Global Aeronautical Engineering and Satellite Technology (GAST)* (pp. 1–5). IEEE. https://doi.org/10.1109/GAST60528.2024.10520774

An, Y., Yu, F. R., He, Y., Li, J., Chen, J., & Leung, V. C. M. (2024). A novel Internet of Things web attack detection architecture based on the combination of symbolism and connectionism AI. *IEEE Internet of Things Journal*, *11*(11), 19823–19837. https://doi.org/10.1109/JIOT.2024.3369852

Arbel, Y., Tokson, M., & Lin, A. (2024). Systemic regulation of artificial intelligence. *Arizona State Law Journal*, *56*(2), 545–619. https://heinonline.org/HOL/LandingPage?handle=hein.journals/arzjl56&div=17&id=&page=

Atif, D. (2024). Enhancing long-term GDP forecasting with advanced hybrid models: A comparative study of ARIMA-LSTM and ARIMA-TCN with dense regression. *Computational Economics*. Advance online publication. https://doi.org/10.1007/s10614-024-10683-5

Auerbach, P., & Green, F. (2024). Reformulating the critique of human capital theory. *Journal of Economic Surveys*. Advance online publication. https://doi.org/10.1111/joes.12675

BCIT. (2026). *Power engineering*. https://www.bcit.ca/programs/power-engineering/

Beard, E., Marsden, J., Brown, J., Tombor, I., Stapleton, J., Michie, S., & West, R. (2019). Understanding and using time series analyses in addiction research. *Addiction*, *114*(10), 1866–1884. https://doi.org/10.1111/add.14643

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5), 9–49. https://doi.org/10.1086/258724

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y. N., Zhang, Y.-

Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., … Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. *Science*, *384*(6698), 842–845. https://doi.org/10.1126/science.adn0117

Bi, H., & Neethirajan, S. (2024). Satellite data and machine learning for benchmarking methane concentrations in the Canadian dairy industry. *Sustainability*, *16*(23), 10400. https://doi.org/10.3390/su162310400

Boskovic, S. (2017). *The persistence of second- and first-class power engineering distance education students at BCIT: A grounded theory approach* [Master's thesis, Simon Fraser University]. Summit Research Repository. https://summit.sfu.ca/item/17706

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control*. John Wiley & Sons. https://www.wiley.com/en-us/Time+Series+Analysis%3A+Forecasting+and+Control%2C+5th+Edition-p-9781118675021

British Columbia. (2026). *Safety Standards Act* (B.C. Reg. 104/2004). https://www.bclaws.gov.bc.ca/civix/document/id/complete/statreg/17_104_2004

Bruiger, D. (2025). Reflections on the AI alignment problem. *AI & SOCIETY: Journal of Knowledge, Culture and Communication*. Advance online publication. https://doi.org/10.1007/s00146-025-02211-2

Byju, S. (2024, April). Rainfall prediction using SARIMA and OpenAI. In *2024 International Conference on Cognitive Robotics and Intelligent Systems (ICC-ROBINS)* (pp. 519–523). IEEE. https://doi.org/10.1109/ICC-ROBINS60238.2024.10533971

Cárdenas, L. (2024). Why does Okun's law change? Essay in econometric history. *Cliometrica: Journal of Historical Economics and Econometric History*. Advance online publication. https://doi.org/10.1007/s11698-024-00290-7

Cardona-Arenas, C. D. (2025). Exploring unemployment persistence: A probabilistic analysis in 20 OECD countries to understand its social implications. *International Journal of Sociology & Social Policy*, *45*(1/2), 86–105. https://doi.org/10.1108/IJSSP-06-2024-0245

Castaño Camps, E. (2022). *Introduction to time series and forecasting*. Universitat de Barcelona. https://hdl.handle.net/2445/189584

Chemlal, O., & Benomar, W. (2024). The technological impact on employment in Spain between 2023 and 2035. *Forecasting*, *6*(2), 296–325. https://doi.org/10.3390/forecast6020017

Cheng, Z., Xu, H., Wang, Y., & Zhou, H. (2023, September). Time series forecasting based on ARIMA and LSTM models. In *2023 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEC)* (pp. 359–364). IEEE. https://doi.org/10.1109/PEEEC60561.2023.00076

Coşar, E. E., & Yavuz, A. A. (2021). Okun's law under the demographic dynamics of the Turkish labor market. *Central Bank Review*, *21*(2), 59–69. https://doi.org/10.1016/j.cbrev.2021.03.002

Domino. (2026). *What is Scikit-learn?* https://domino.ai/data-science-dictionary/sklearn

Ellis, G. (2010). *Strategies to increase student completion rates in the second-class power engineering program at the British Columbia Institute of Technology* [Master's thesis, Royal Roads University]. Library and Archives Canada.

Gajewski, P., Čule, B., & Rankovic, N. (2023). Unveiling the power of ARIMA, support vector and random forest regressors for the future of the Dutch employment market. *Journal of Theoretical & Applied Electronic Commerce Research*, *18*(3), 1365–1403. https://doi.org/10.3390/jtaer18030069

Getachew, M., Beshah, B., Mulugeta, A., & Kitaw, D. (2024). Application of artificial intelligence to enhance manufacturing quality and zero-defect using CRISP-DM framework. *International Journal of Production Research*, 1–25. https://doi.org/10.1080/00207543.2024.2407919

Gil-Alana, L. A. A., González-Blanch, M. J., & Poza, C. (2025). Labour market mismatches in G7 countries: A fractional integration approach. *Applied Economics*, *57*(7), 736–752. https://doi.org/10.1080/00036846.2024.2305620

Glackin, C. E. W., & Adivar, M. (2023). Using the power of machine learning in sales research: Process and potential. *Journal of Personal Selling & Sales Management*, *43*(3), 178–194. https://doi.org/10.1080/08853134.2022.2128812

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep feedforward networks. In *Deep learning* (pp. 161–217). MIT Press. https://mnassar.github.io/deeplearninghandbook/slides/06_mlp.pdf

Gopu, A., Ramakrishnan, A., Balasubramanian, G., & Srinidhi, K. (2023, April). A comparative study on forest fire prediction using ARIMA, SARIMA, LSTM, and GRU methods. In *2023 IEEE International Conference on Contemporary Computing and Communications (InC4)* (Vol. 1, pp. 1–5). IEEE. https://doi.org/10.1109/InC457730.2023.10262965

Government of Canada. (2025). *Job prospects: Power engineer in Canada*. https://www.jobbank.gc.ca/marketreport/outlook-occupation/7857/ca

Hanna, A., Nye, C. D., Samo, A., Chu, C., Hoff, K. A., Rounds, J., & Oswald, F. L. (2024). Interests of the future: An integrative review and research agenda for an automated world of work. *Journal of Vocational Behavior*, *152*, Article 104012. https://doi.org/10.1016/j.jvb.2024.104012

Hau, H. (2007). *Dealing with the shortage of power engineers in NE Alberta* [Master's thesis]. ProQuest Dissertations and Theses database.

Huang, W., Wang, D., Pi, X., & Hewlin, P. F. (2022). Does coworkers' upward mobility affect employees' turnover intention? The roles of perceived employability and prior job similarity. *International Journal of Human Resource Management*, *33*(12), 2442–2472. https://doi.org/10.1080/09585192.2020.1861058

Huruta, A. D. (2024). Predicting the unemployment rate using autoregressive integrated moving average. *Cogent Business & Management*, *11*(1), 1–21.

https://doi.org/10.1080/23311975.2023.2293305

Ibrahim, N., Ahmad, N., Jan, N. A. M., Zainudin, Z., Jamil, N. S., & Azlan, A. (2024, September). Comparative analysis of ARIMA and LSTM approaches for monthly river flow forecasting in Terengganu. In *2024 5th International Conference on Artificial Intelligence and Data Sciences (AiDAS)* (pp. 1–6). IEEE. https://doi.org/10.1109/AiDAS63860.2024.10730554

Institute of Power Engineers. (2026). *Home*. https://nipe.ca/

Jain, R., Khairnar, A. S., & Karthikeyan, S. P. (2024, October). Forecasting Indian conventional generation: A comparative analysis of ARIMA and SARIMA models for improved energy demand prediction. In *2024 International Conference on Sustainable Energy: Energy Transition and Net-Zero Climate Future (ICUE)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICUE63019.2024.10795531

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). *An introduction to statistical learning: With applications in Python*. Springer. https://www.statlearning.com/

Jin, M. (2024, November). Research on cargo volume prediction in logistics sorting centers based on ARIMA-LSTM forecasting algorithm. In *2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM)* (pp. 644–648). IEEE. https://doi.org/10.1109/ISPCEM64498.2024.00116

Johnston, L., Patel, V., Cui, Y., & Balaprakash, P. (2025). Revisiting the problem of learning long-term dependencies in recurrent neural networks. *Neural Networks*, *183*, 106887. https://doi.org/10.1016/j.neunet.2024.106887

Kapadia, N., & Mehta, R. (2025). Energy efficient waste collection vehicle routing system using time series prediction. *Multimedia Tools and Applications*, *84*(9), 6157–6186. https://doi.org/10.1007/s11042-024-18894-6

Keras. (2026). *Developer guides*. https://keras.io/

Khalaf, C., Michaud, G., & Jolley, G. J. (2023). Predicting declining and growing occupations using supervised machine learning. *Journal of Computational Social Science*, *6*(2), 757–780. https://doi.org/10.1007/s42001-023-00211-0

Kolassa, S. (2025). Know your errors! *Foresight: The International Journal of Applied Forecasting*, *78*, 31–32. https://ideas.repec.org/a/for/ijafaa/y2025i78p31-32.html

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, *15*(8), 255. https://doi.org/10.3390/fi15080255

Krstic, Z. (2024). Economic theory and artificial intelligence: A cross-model perspective on labour market dynamics. *Croatian Regional Development Journal*, *5*(2), 52–75. https://doi.org/10.2478/crdj-2024-0008

Kuhn, T. S. (1997). *The structure of scientific revolutions* (Vol. 962). University of Chicago Press.

Lalé, E. (2024). *Assessing labor market conditions in Canada with public-use microdata* (IZA Discussion Paper No. 16722). SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4692431

Li, W., & Li, T. (2025). Comparison of deep learning models for predictive maintenance in industrial manufacturing systems using sensor data. *Scientific Reports*, *15*(1), 1–19. https://doi.org/10.1038/s41598-025-08515-z

Lin, W. T., Chen, Y. H., & Jhang, S.-S. (2023). The effects of unemployment and inflation rates on the business value of information technology and economic performance: The partial adjustment valuation approaches. *Asia Pacific Management Review*, *28*(4), 371–389. https://doi.org/10.1016/j.apmrv.2022.12.005

Lutchman, C. (2008). *Leadership impact on turnover among power engineers in the oil sands of Alberta* (Doctoral dissertation). ProQuest Dissertations and Theses Global.

Marinaș, L. E., Păun, C. V., Diaconescu, M., & Smirna, T. G. (2024). Artificial intelligence readiness and employment: A global panel analysis. *Economic Computation & Economic Cybernetics Studies & Research*, *58*(4), 57–74. https://doi.org/10.24818/18423264/58.4.24.04

Mavrodi, A. G., Kolias, G., & Karamanis, K. (2024). An empirical analysis of Okun's law at a regional level: Evidence from the NUTS 2 European regions. *Applied Economics*, *56*(56), 7600–7620. https://doi.org/10.1080/00036846.2023.2288043

Mishra, R. K., Ansari, A. R., Jothi, J. A. A., & Mishra, V. (2024). Analysis of criminal landscape by utilizing statistical analysis and deep learning techniques. *Journal of Applied Security Research*, *19*(4), 560–585. https://doi.org/10.1080/19361610.2024.2314392

Mohamed, A. A. (2024). Economic growth and unemployment nexus: Empirical test of Okun's law in Somalia. *Journal of Economic Structures*, *13*(1), 1–12. https://doi.org/10.1186/s40008-024-00337-y

Mullen, C. R. (2017). *Factors influencing Canadian power engineers' decision to pursue advanced certification* (Order No. 10602638) [Doctoral dissertation]. ProQuest Dissertations & Theses Global.

Orozco, C. J. M., Sierra, S. L. P., & Vidal, P. (2024). Labor market forecasting in unprecedented times: A machine learning approach. *Bulletin of Economic Research*, *76*(4), 893–915. https://doi.org/10.1111/boer.12451

Pimenta, M. L., da Silva, A. L., & Tate, W. L. (2016). Characteristics of cross-functional integration processes. *International Journal of Logistics Management*, *27*(2), 570–594. https://doi.org/10.1108/IJLM-01-2014-0010

Prorok, M., & Takacs, I. (2024, May). The impacts of artificial intelligence and knowledge-based systems on corporate decision support. In *2024 IEEE 18th International Symposium on Applied Computational Intelligence and Informatics (SACI)* (pp. 65–70). IEEE. https://doi.org/10.1109/SACI60582.2024.10619820

Rahhal, I., Kassou, I., & Ghogho, M. (2024). Data science for job market analysis: A survey on applications and techniques. *Expert Systems with Applications*, *251*, 124101. https://doi.org/10.1016/j.eswa.2024.124101

Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, *98*(5, Part 2), S71–S102. https://web.stanford.edu/~klenow/Romer_1990.pdf

Sangeetha, S., Mathivanan, S. K., Rajadurai, H., Cho, J., & Easwaramoorthy, S. V. (2024). A multi-modal geospatial–temporal LSTM based deep learning framework for predictive modeling of urban mobility patterns. *Scientific Reports*, *14*(1), 1–19. https://doi.org/10.1038/s41598-024-74237-3

Sasaki, H. (2024). Effects of product and process innovation on growth and unemployment. *SN Business & Economics*, *4*(11). https://doi.org/10.1007/s43546-024-00733-3

Schumpeter, J. A. (2021). *The theory of economic development*. Routledge. https://doi.org/10.4324/9781003146766

Selmy, H. A., Mohamed, H. K., & Medhat, W. (2024). A predictive analytics framework for sensor data using time series and deep learning techniques. *Neural Computing & Applications*, *36*(11), 6119–6132. https://doi.org/10.1007/s00521-023-09398-9

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 1394–1401). IEEE. https://doi.org/10.1109/ICMLA.2018.00227

Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. *IEEE Access*, *10*, 124715–124727. https://doi.org/10.1109/ACCESS.2022.3224938

Solow, R. M. (1957). Technical change and the aggregate production function. *The Review of Economics and Statistics*, *39*(3), 312–320. https://doi.org/10.2307/1926047

Statistics Canada. (2025). *Job vacancy and wage survey (JVWS)*. https://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5217

Statistics Canada. (2025). *Labour force survey (LFS)*. https://www.statcan.gc.ca/en/survey/household/3701

Statistics Canada. (2024). *North American Industry Classification System (NAICS)*. https://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVD&TVD=1369825

Statistics Canada. (2025). *National Occupational Classification (NOC)*. https://noc.esdc.gc.ca/
Statistics Canada. (2025). *Section 3: Terms and definitions*. https://www150.statcan.gc.ca/n1/pub/72-210-g/2013001/part-partie3-eng.htm

Statistics Canada. (2025). *Standards, data sources and methods*. https://www.statcan.gc.ca/en/concepts/index

Statistics Canada. (2025). https://www.statcan.gc.ca/en/start

Suman, S., Kaur, S. J., Sharma, A., & Kumar, S. (2024, February). Machine learning-based system for admission and jobs prediction in engineering and technology sector. In *2024*

IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) (Vol. 5, pp. 463–468). IEEE. https://doi.org/10.1109/IC2PCT60090.2024.10486533

Sun, Y., & Dimiski, A. (2024). Exploring inflation dynamics in Canada: A threshold vector autoregressive approach. *The Journal of Economic Asymmetries*, *30*, e00364. https://doi.org/10.1016/j.jeca.2024.e00364

Sun, W., Wang, K., Wang, M., & Lin, Q. (2025). L21 norm regularization ELM with p-Huber loss function for multi-target regression. *IAENG International Journal of Applied Mathematics*, *55*(8), 2570–2580. https://www.iaeng.org/IJAM/issues_v55/issue_8/IJAM_55_8_28.pdf

TensorFlow. (2026). *TensorFlow core*. https://www.tensorflow.org/

Townsend, D. M., Hunt, R. A., Rady, J., Manocha, P., & Jin, J. H. (2024). Are the futures computable? Knightian uncertainty and artificial intelligence. *Academy of Management Review*, *50*(2), 415–440. https://doi.org/10.5465/amr.2022.0237

TSBC. (2026). *Technical Safety BC*. https://www.technicalsafetybc.ca/

Ullah, M. A., Mim, A. S., Hasan, M. N., & Sadik, M. R. (2024, May). Deep learning based forecasting models of dengue outbreak in Bangladesh: Comparative analysis of LSTM, RNN, and GRU models using multivariate variables with a two-decade dataset. In *2024 International Conference on Smart Systems for Applications in Electrical Sciences (ICSSES)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICSSES62373.2024.10561382

Waqas, M., & Humphries, U. W. (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*, *13*, 102946. https://doi.org/10.1016/j.mex.2024.102946

Weiss, Y. (2015). Gary Becker on human capital. *Journal of Demographic Economics*, *81*(1), 27–31. doi:10.1017/dem.2014.4

Wharton University of Pennsylvania. (2026). *WRDS Wharton research data services*. https://wrds-www.wharton.upenn.edu/pages/about/what-wrds/

Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A comparison between ARIMA, LSTM, and GRU for time series forecasting. *Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence*, 49–55. https://doi.org/10.1145/3377713.3377722

Zeng, L., Ling, L., Zhang, D., & Jiang, W. (2023). Optimal forecast combination based on PSOCS approach for daily agricultural future prices forecasting. *Applied Soft Computing*, *132*, 109833. https://doi.org/10.1016/j.asoc.2022.109833

Zhu, L., Gao, J., Zhu, C., & Deng, F. (2025). Short-term power load forecasting based on spatial-temporal dynamic graph and multi-scale Transformer. *Journal of Computational Design & Engineering*, *12*(2), 92–111. https://doi.org/10.1093/jcde/qwaf013

Zhu, X., Xu, J., Fu, Z., Damarla, S. K., Wang, P., & Hao, K. (2025). Novel dynamic data-driven modeling based on feature enhancement with derivative memory LSTM for complex industrial process. *Neurocomputing*, *626*, 129619. https://doi.org/10.1016/j.neucom.2025.129619