

Optimizing Time Series Forecasting for AI Job Market Dynamics: A Comparative Study of SARIMA and Prophet with Cross-Validation

Slamet Widodo^{1,*}, Budi Santoso²

^{1,2}Magister of Computer Sciences, Amikom Purwokerto University, Indonesia

(Received June 1, 2025; Revised October 5, 2025; Accepted January 20, 2026; Available online March 25, 2026)

Abstract

The rapid growth of Artificial Intelligence (AI) has led to significant shifts in the job market, making accurate forecasting essential for workforce planning and policy-making. This study aims to compare the effectiveness of two time series forecasting models, SARIMA and Prophet, in predicting AI job market trends. The primary focus was to evaluate their performance using RMSE and Time Series Cross-Validation (TSCV). The SARIMA model outperformed Prophet, particularly when capturing seasonal patterns, achieving an RMSE of 83.42, compared to Prophet's 136.01. The results demonstrate that SARIMA's parameter tuning and seasonality handling contribute to more reliable forecasts, while Prophet's flexibility offers advantages in non-linear data scenarios. The study provides valuable insights for stakeholders in adapting workforce strategies, particularly in AI-driven sectors, by recommending SARIMA for seasonal trends and Prophet for volatile, non-linear data.

Keywords: SARIMA, Prophet, AI Job Market, Time Series Forecasting, Model Comparison

1. Introduction

The rapid advancement of Artificial Intelligence (AI) is reshaping various sectors of the economy, significantly impacting job market dynamics. As AI becomes more integrated into industries, its influence on employment trends becomes increasingly critical to understand. The need for accurate forecasting of job market trends influenced by AI is essential for policymakers, business leaders, and workers, who must navigate the complexities introduced by this technological shift.

AI's impact on the labor market is twofold. While AI automates repetitive tasks, leading to job displacement in sectors vulnerable to automation, it simultaneously creates new opportunities in emerging technology fields. For example, AI has spurred job growth in technology-related sectors, while also driving job losses in lower-skilled roles [1], [2]. This dual impact highlights the need for the workforce to adapt to shifting skill demands as AI evolves. However, as traditional education systems may not adequately prepare individuals for these new roles, there is a growing need for retraining and upskilling initiatives [3].

Accurately forecasting the implications of AI on the labor market is crucial for both public policy and workforce planning. While traditional forecasting methods like ARIMA and exponential smoothing are commonly used, they often fail to capture the complexities of labor market trends affected by AI, including seasonality, outliers, and non-stationarity. These factors make it difficult to predict future job trends accurately, which poses a challenge for organizations and policymakers.

This study seeks to address the challenge of forecasting AI job openings by comparing two popular time series models: Seasonal Auto-Regressive Integrated Moving Average (SARIMA) and Facebook Prophet. Both models are widely used in forecasting applications, and each has strengths in handling seasonality and non-linear trends. The primary objective is to determine which model performs better in predicting job openings in the context of AI's impact on the labor market.

*Corresponding author: Slamet Widodo (swidodo009@gmail.com)

DOI: <https://doi.org/10.47738/ijjis.v9i2.296>

This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights

To evaluate the performance of both models, the study will focus on Root Mean Squared Error (RMSE), a key metric used to measure the accuracy of the forecasts. RMSE will be used to assess which model best captures the underlying trends in the job market data.

To ensure the robustness of the evaluation, Time Series Cross-Validation will be employed. This method involves splitting the time series data into multiple segments, training the model on one segment, and testing it on subsequent segments. This technique mimics real-world forecasting scenarios and provides a more reliable assessment of how well each model performs when making future predictions based on past data. By using Time Series Cross-Validation, the study will ensure that the models' forecasts are tested under various time periods, accounting for changing patterns in the data.

2. Literature Review

2.1. The Dual Impact of AI on Job Market Dynamics: Displacement and Creation

The impact of Artificial Intelligence (AI) on the labor market is multifaceted, with both the displacement of traditional jobs and the creation of new roles being evident. AI technologies, particularly those automating repetitive tasks, have led to the displacement of low-skilled jobs in sectors like manufacturing and finance [4], [5]. These roles, characterized by routine and manual tasks, are increasingly vulnerable to automation, a phenomenon referred to as the "replacement effect" [6]. For example, manufacturing and service sector jobs that rely heavily on repetition are particularly susceptible to being replaced by AI technologies [7], [8]. However, alongside this displacement, AI also enhances productivity and alters required skill sets, thereby favoring the high-skilled labor market [5], [9], [10].

Conversely, AI is a significant driver of job creation in emerging fields, particularly in tech-centric sectors such as data analysis, AI engineering, and AI management [10]. As organizations increasingly adopt AI technologies, the demand for professionals with expertise in AI development, data analysis, and ethical governance has surged [11]. This shift highlights the transformative role of AI in creating new job categories that require advanced technical skills, marking a clear shift in workforce demand [12]. As AI continues to shape the labor market, these new roles underscore the growing importance of highly skilled professionals in managing and advancing AI technologies.

The dual impact of AI on job displacement and creation calls for proactive strategies from policymakers to manage the workforce transition. While AI creates new opportunities, particularly in high-tech and data-driven industries, it also poses risks to sectors reliant on routine tasks. Policymakers must implement strategies that balance the benefits of AI with measures to mitigate job losses, ensuring that workers are equipped with the skills required for emerging roles [13]. In conclusion, understanding AI's dual influence on the job market is crucial for preparing for both the challenges it presents and the opportunities it offers, necessitating adaptive strategies across various sectors [13], [14].

2.2. Overview of SARIMA and Prophet Models

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is an extension of the ARIMA model, specifically designed to capture seasonal variations in time series data. SARIMA combines three key components: Autoregression (AR), Differencing (I), and Moving Averages (MA), along with seasonal terms to handle periodic fluctuations. This makes it particularly useful for datasets exhibiting both trends and seasonal patterns, such as in economic forecasting or environmental studies [15], [16]. SARIMA is represented as SARIMA (p,d,q) (P,D,Q)_s, where the parameters p, d, q corresponds to the autoregressive terms, differencing, and moving averages for non-seasonal data, while P, D, Q are the seasonal components, and s represents the length of the seasonal cycle [16], [17]. This structure enables SARIMA to effectively model seasonal data and makes reliable predictions across various fields.

One of SARIMA's main advantages is its ability to adapt to different seasonal patterns, making it a versatile tool for time series analysis. It has been widely used in forecasting applications, such as predicting inflation rates, demand cycles in manufacturing, and environmental parameters like rainfall [18]. Its robustness and statistical foundation make it appealing to both researchers and practitioners in fields ranging from economics to meteorology. However, SARIMA requires careful tuning of its parameters to achieve optimal forecasting accuracy, and it may not perform well in non-linear or highly irregular datasets.

On the other hand, Prophet, an open-source forecasting tool developed by Facebook, is designed to handle time series data that exhibits non-linear trends, seasonal effects, and occasional missing values or outliers. Prophet is particularly effective in real-world applications where traditional models may struggle with data anomalies or irregular trends [19]. Prophet's core feature is its additive model structure, which separates trend, seasonality, and holiday effects, allowing for clear interpretations of underlying patterns. Prophet uses a piecewise linear or logistic growth curve to model trends, making it adept at capturing non-linear patterns in the data [20], [21]. This flexibility allows users to incorporate prior knowledge about the data, such as seasonal effects due to holidays or other special events, resulting in more accurate forecasts.

Prophet's ease of use is one of its key strengths. Unlike SARIMA, which requires detailed parameter tuning, Prophet is user-friendly and requires minimal adjustments, making it accessible even for those without extensive data science expertise [19]. Its versatility has led to its widespread use in various domains, such as forecasting air quality parameters [22], predicting retail demand, and analyzing public health trends like suicide rates and disease outbreaks [23]. The model's robustness in handling missing data and its adaptability to sparse datasets further enhance its appeal for real-world forecasting challenges.

Both SARIMA and Prophet offer valuable tools for time series forecasting, each with its own strengths. SARIMA excels in handling seasonal data with clear patterns and is widely used in structured datasets, while Prophet shines in its ability to handle non-linear trends and irregularities in real-world data with minimal tuning. By integrating both models into forecasting strategies, researchers and practitioners can choose the most suitable approach based on the characteristics of the data and the forecasting requirements.

2.3. Forecasting Accuracy Metrics and Model Validation

Accurate forecasting relies heavily on evaluating model performance through established metrics. The most commonly used metrics in time series forecasting include RMSE, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), each offering unique insights into model performance. RMSE measures the square root of the average of squared differences between predicted and observed values, emphasizing larger errors, making it sensitive to outliers. A lower RMSE indicates better forecasting accuracy and is widely used in energy and health forecasting [24]. MAE, on the other hand, averages the absolute differences between forecasted and actual values, treating all errors equally, and is often used in domains like COVID-19 forecasting [25]. MAPE expresses accuracy as a percentage, which is particularly useful for comparing performance across datasets of different scales, such as in financial forecasting [26], [27].

The choice of evaluation metric depends on the context of the data and forecasting objectives. RMSE penalizes larger errors, making it suitable for applications where outliers should be emphasized. MAE provides a straightforward and interpretable measure of error magnitude, while MAPE offers a relative measure of accuracy, particularly useful when comparing models across different datasets. Each of these metrics provides critical insights into how well a model performs, and using them together allows for a comprehensive assessment of forecasting accuracy.

Model validation plays a crucial role in ensuring the reliability of forecasting models, particularly in time series analysis where underlying patterns may shift over time. Time Series Cross Validation (TSCV) is a technique specifically adapted to time series data, which respects the sequential nature of the data by using past observations to predict future values. This technique typically involves a rolling window approach, where the model is trained on an expanding set of data and validated on future points, ensuring that the temporal structure is maintained [28], [29]. TSCV helps to minimize overfitting and ensures that the model generalizes well to different time periods, making it an essential tool for time series forecasting [30].

In addition to TSCV, nested cross-validation can further mitigate model evaluation bias. This two-layer technique tunes the model within the inner loop and assesses its predictive power in the outer loop, providing a more robust evaluation. This approach is particularly effective for non-stationary or dynamic datasets, ensuring the validity of performance assessments [31]. Using rigorous validation methods like TSCV is essential not only for assessing the model's performance on training data but also for ensuring that it maintains predictive integrity when applied to unseen future

data. By combining accurate forecasting metrics with these validation techniques, practitioners can ensure the reliability and accuracy of their models.

3. Methodology

3.1. Data Collection and Preprocessing

The dataset used for this study consists of historical data on AI job market trends, specifically focused on job openings related to artificial intelligence across various industries. The dataset includes information on job titles, industry sectors, geographical location, and other relevant attributes such as required skills, qualifications, and salary ranges. This data has been collected from reliable sources such as job boards, company databases, and labor market surveys, spanning multiple years to capture significant trends and fluctuations in AI job demand.

To ensure the data is suitable for time series modeling, a logarithmic transformation (`np.log1p()`) was applied to stabilize variance and normalize the distribution. The original data exhibited significant skewness and extreme outliers, which could negatively affect the performance of forecasting models. The logarithmic transformation helps reduce the impact of large values, allowing the time series data to better meet the assumptions of stationarity required for models like SARIMA and Prophet.

Missing data and outliers are common challenges in time series datasets. For missing values, interpolation techniques were employed to estimate missing entries, ensuring the continuity of the dataset. Outliers were identified using statistical methods such as Z-scores and the Interquartile Range (IQR). Extreme outliers that could distort model predictions were either winsorized or transformed, ensuring they did not disproportionately influence the results. This preprocessing ensures the dataset is clean and suitable for effective modeling, improving the accuracy and reliability of the forecasting models.

3.2. Time Series Forecasting Models

The SARIMA model is an extension of the ARIMA model that specifically incorporates seasonal components to account for periodic fluctuations in time series data. SARIMA consists of three main parts: AR, I, and MA, and it further introduces seasonal terms to handle seasonality in the data. The SARIMA model is represented as:

$$\text{SARIMA}(p, d, q)(P, D, Q)_s \quad (1)$$

Where:

p is the number of autoregressive terms.

d is the degree of differencing applied to make the series stationary.

q is the number of moving average terms.

P , D , and Q are the seasonal components corresponding to autoregressive, differencing, and moving average terms at the seasonal level.

s represents the length of the seasonal cycle, such as 12 for monthly data with yearly seasonality.

The key challenge in using SARIMA is parameter tuning, where the optimal values of p , d , q , P , D , Q , and s are determined. These values can be found using techniques like grid search or automatic selection methods such as `auto_arima`, which evaluates combinations of parameters based on minimizing performance metrics like the Akaike Information Criterion (AIC) or MSE.

SARIMA is built on the assumption that the data is stationary after differencing and that there is a linear relationship between past observations and the forecasted values. However, SARIMA has limitations, especially when handling non-linear relationships or sudden shifts in the data. For example, when forecasting data influenced by disruptive events, such as sudden changes in AI job market trends, SARIMA might struggle to capture these non-linear patterns effectively. Another limitation of SARIMA is that it requires extensive manual tuning of parameters, which can be computationally expensive and time-consuming, particularly for large datasets.

In contrast, Prophet, developed by Facebook, is a forecasting model designed to handle time series data with non-linear trends, seasonal effects, and other irregularities like missing data and outliers. Prophet's core strength lies in its additive model structure, which allows for clear separation of the underlying trend, seasonality, and holiday effects. The model uses a piecewise linear or logistic growth curve to model the trend component, which makes it adept at capturing non-linear behaviors that are common in real-world datasets. The general form of the Prophet model is given by:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (2)$$

Where:

$y(t)$ is the observed value at time t .

$g(t)$ is the trend function, which can be piecewise linear or logistic.

$s(t)$ is the seasonal component.

$h(t)$ represents the holiday effects.

ϵ_t is the error term, capturing the residual noise.

Prophet is highly flexible, allowing users to adjust parameters such as the seasonalities and holidays, making it suitable for datasets influenced by external events or special occasions. Its ability to handle missing data is particularly noteworthy, as it automatically accounts for gaps in the time series without requiring imputation or interpolation. Prophet is also known for its minimal tuning requirements. Unlike SARIMA, which often requires manual adjustment of many parameters, Prophet can be used with default settings and still perform well in most cases.

One of Prophet's key advantages is its flexibility in handling irregular data, such as missing values or outliers. In cases where AI job market data may have gaps or anomalies due to sporadic data collection, Prophet can still produce reliable forecasts by adjusting for these irregularities. This makes Prophet especially useful in real-world forecasting scenarios where data quality is not always consistent or complete.

3.3. Time Series Cross-Validation

TSCV is an essential technique for evaluating the performance of forecasting models, especially in time series data where observations are dependent on previous time points. Unlike traditional cross-validation methods, which randomly split the data into training and testing sets, rolling forecast origin cross-validation ensures that the temporal order of the data is preserved. This method works by progressively training the model on a growing set of historical data, while testing it on the subsequent time points. The forecast origin shifts forward in time with each iteration, ensuring that each test set consists of unseen data, mimicking the real-world scenario of making future predictions based on past observations.

The rolling forecast origin cross-validation process can be described as follows: Initially, a model is trained on a small initial dataset and evaluated on the following period. Then, the training set is expanded by including the next time point, and the model is retrained and tested again. This process continues until all the data points are used for testing, providing a series of predictions that span the entire dataset. This method ensures that the model is evaluated on data it has not seen during training, helping to assess its predictive performance over time.

The importance of cross-validation, particularly for time series data, lies in its ability to assess model robustness across multiple forecast horizons. Since time series data often exhibits trends, seasonality, and other temporal dependencies, it is crucial to evaluate how well a model generalizes to different points in time, especially under varying conditions. TSCV helps identify how well the model can predict future data by testing its performance at multiple points across the forecast horizon. By using multiple forecast horizons, we can gauge the model's ability to adapt to changes in the underlying patterns of the data, such as sudden shifts in trends or seasonal variations. This method also helps detect potential issues like overfitting, where a model may perform well on historical data but fail to predict future values effectively.

In summary, TSCV is an indispensable tool for assessing the stability and reliability of forecasting models. By evaluating model performance across different forecast horizons, it provides a more comprehensive understanding of

how well a model is likely to perform when applied to future, unseen data. This process is essential for ensuring that the forecasting model not only fits the historical data well but is also capable of making accurate predictions in real-world applications.

3.5. Performance Evaluation

In forecasting, RMSE is a widely used metric for evaluating the accuracy of predictive models. RMSE measures the square root of the average squared differences between the observed actual outcomes and the predicted values. The formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (3)$$

y_t represents the actual values at time t , \hat{y}_t represents the predicted values at time t , n is the number of observations.

RMSE is particularly useful for comparing models, as it gives a clear indication of the magnitude of prediction errors. A lower RMSE indicates better model performance, as it shows that the model's predictions are closer to the actual values. RMSE also has the advantage of penalizing larger errors due to the squaring of the differences, making it sensitive to significant deviations in the predictions. This makes it especially valuable when large forecasting errors are undesirable, as is often the case in applications like AI job market forecasting.

In this study, RMSE will be used as the primary metric for evaluating the performance of the SARIMA and Prophet models. By calculating the RMSE for both models, we can compare their ability to accurately predict AI job market trends. The SARIMA model is a traditional statistical approach that is highly effective in modeling seasonal data, but its performance depends heavily on the tuning of its parameters (p , d , q , P , D , Q , s). On the other hand, the Prophet model is designed to handle non-linear trends and irregularities in the data, such as missing values and outliers, with minimal tuning required.

Through RMSE, we will assess how each model performs on the same dataset, comparing their predictive accuracy. A model with a lower RMSE will be considered superior in terms of forecasting accuracy. Additionally, RMSE allows us to determine which model best captures the underlying patterns in the AI job market data, including trends, seasonality, and irregular fluctuations. This comparison will provide valuable insights into which model, SARIMA or Prophet, is more suitable for forecasting AI job openings and meeting the challenges posed by the dynamic nature of the job market.

4. Results

4.1. SARIMA Model Performance

The performance of the initial SARIMA model was evaluated by applying a logarithmic transformation (`np.log1p()`) to the AI job openings data in order to stabilize variance and normalize the distribution. The `auto_arima` function was used to identify the best-fit SARIMA model, resulting in a model configuration of SARIMA (0,1,1) (0,0,0) [12], which was then evaluated on the original test set (after inverse transformation). The RMSE for this initial model was found to be 94.65, indicating a moderate level of forecast accuracy. [Figure 1](#) illustrates the forecasted values from the SARIMA model in comparison with the actual test data. The plot shows the initial forecast's ability to capture trends in the transformed data, with the model's predictions following the general pattern of the actual values, though some discrepancies are evident, especially as the forecast period progresses.

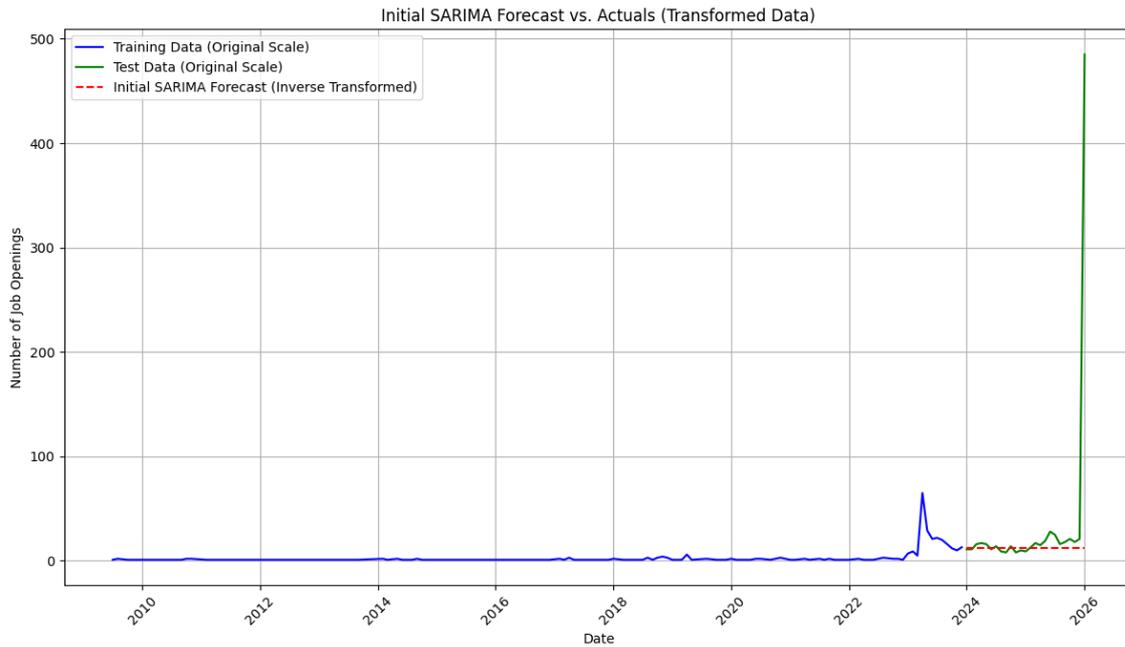


Figure 1. Initial SARIMA Forecast vs. Actuals (Transformed Data)

Next, the residuals of the initial SARIMA model were analyzed to assess how well the model fit the data. The Ljung-Box test conducted on the residuals produced high p-values (e.g., 0.611 for lag 10, 0.930 for lag 20), which indicated that the residuals were white noise, and the model had adequately captured the underlying patterns in the data. To visualize this, Figure 2 shows the residuals plotted against time. The plot demonstrates that the residuals are randomly distributed around zero, suggesting that there is no significant autocorrelation, a sign that the model is well-calibrated and has adequately accounted for the data's structure.

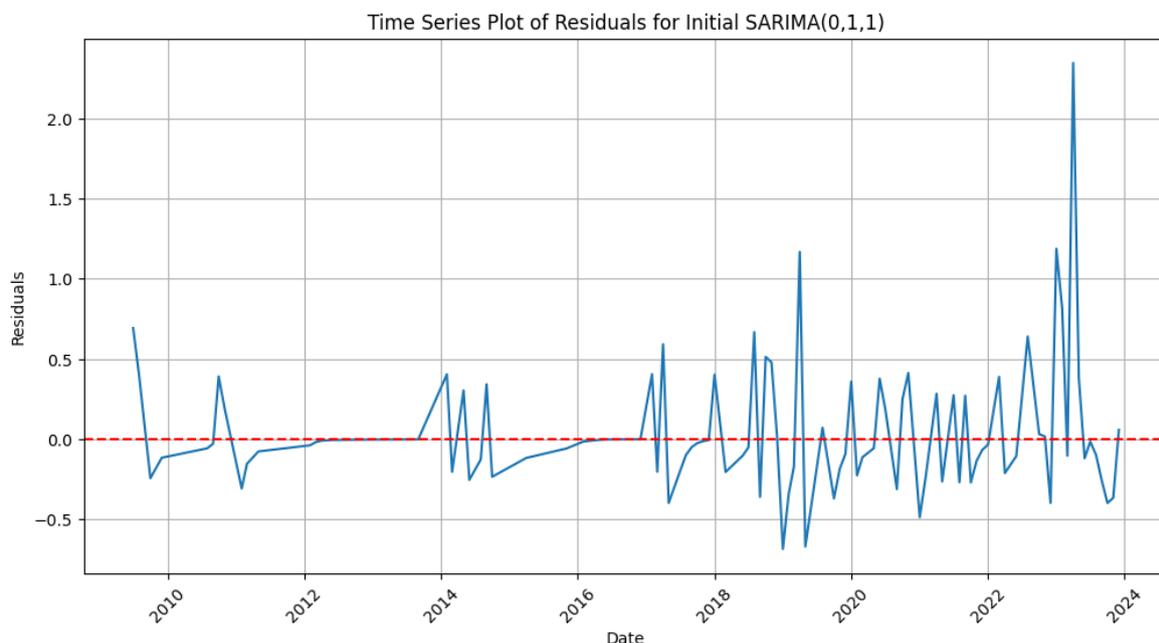


Figure 2. Time Series Plot of Residuals for Initial SARIMA (0,1,1)

An exhaustive grid search was then conducted to optimize the SARIMA model's parameters, which identified two key configurations. The model with the lowest AIC was SARIMA (2,1,0) (1,0,0) [12], with an AIC of 102.947 and an RMSE of 94.503. Meanwhile, the model with the lowest RMSE was SARIMA (0,2,2) (1,1,1) [12], with an RMSE of 83.424 and an AIC of 114.599. Given its superior predictive accuracy, the SARIMA (0,2,2) (1,1,1) [12] model was

chosen for further evaluation. Figure 3 shows the optimized model's performance when applied to an 80/20 train-test split. The optimized SARIMA model's forecast closely matches the actual data, improving on the initial SARIMA model's predictions.

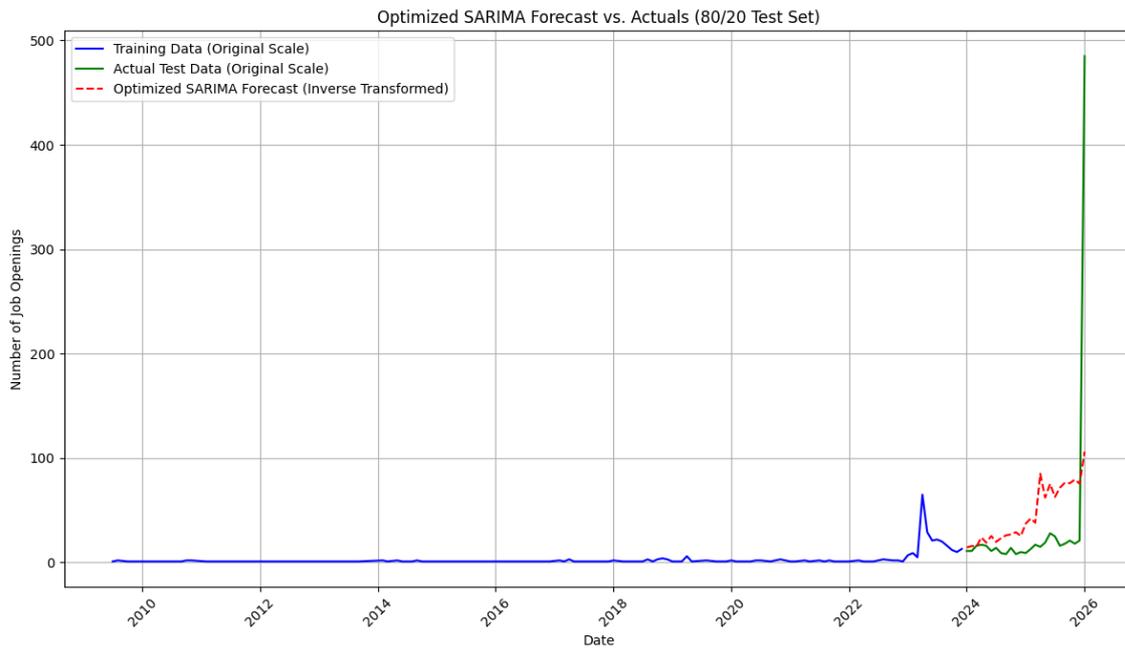


Figure 3. Optimized SARIMA Forecast vs. Actuals (80/20 Test Set)

When the SARIMA (0,2,2) (1,1,1) [12] model was tested on a larger training dataset (approximately 90% of the total data), its performance slightly deteriorated, with the RMSE increasing to 134.6174. This suggests that the model's accuracy can vary significantly depending on the train-test split, highlighting the importance of data characteristics and the potential for overfitting when using larger training datasets. Figure 4 compares the optimized SARIMA model's forecast to the actual data on the 90/10 split. The model performs well on the training data, but its performance decreases when applied to a smaller test set, underlining the variability in forecasting accuracy.

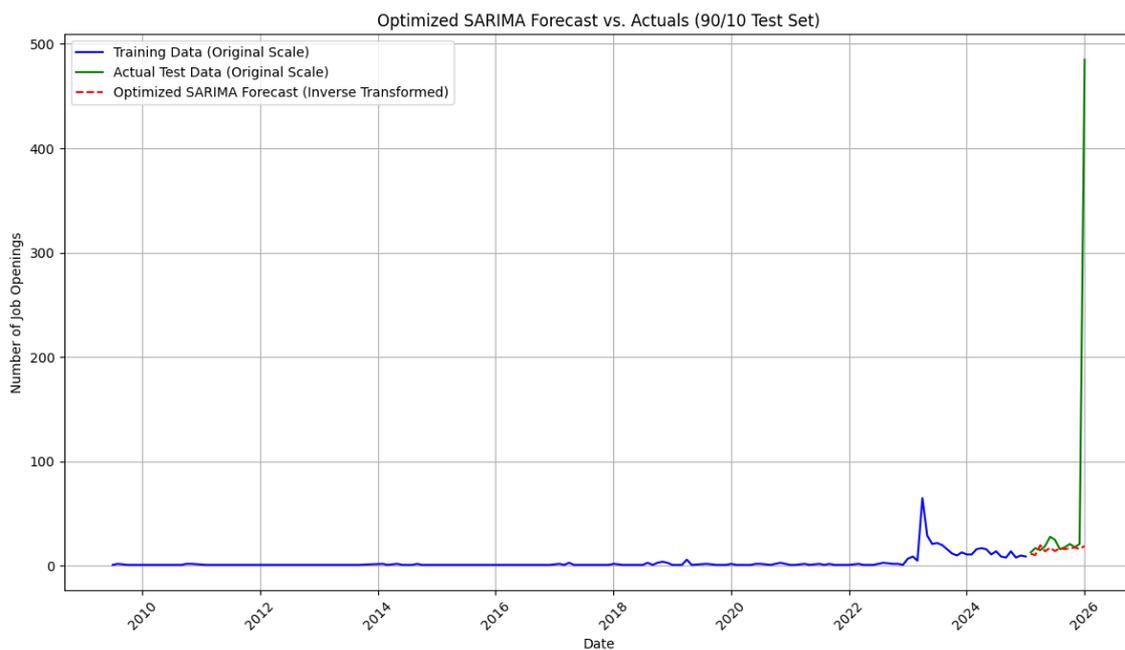


Figure 4. Optimized SARIMA Forecast vs. Actuals (90/10 Test Set)

In addition to optimizing SARIMA, an attempt was made to improve the model by incorporating an exogenous variable (tech_index), which was expected to account for external factors influencing the AI job market. However, this did not significantly improve forecasting accuracy, with the RMSE remaining at 134.64. The exogenous variable's coefficient had a high p-value of 0.952, suggesting that it was not statistically significant in this context. Figure 5 shows the SARIMAX model's forecast, which is similar to the SARIMA model's performance, demonstrating the limited impact of the additional variable on forecast accuracy.

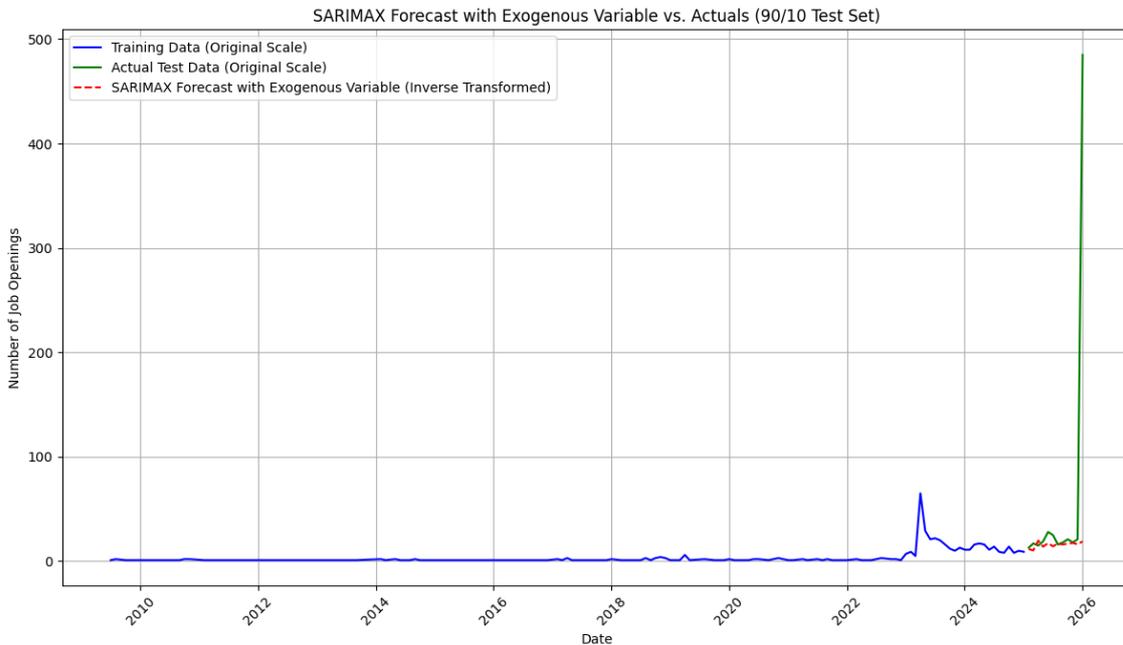


Figure 5. SARIMAX Forecast with Exogenous Variable vs. Actuals (90/10 Test Set)

Finally, the comparison of SARIMA model RMSEs across different configurations is shown in Figure 6. This bar chart compares the RMSE values for the initial SARIMA model, the SARIMA model with the lowest AIC, the SARIMA model with the lowest RMSE, and the optimized model on the 90/10 test set. The chart highlights that while the model with the lowest AIC performed well, the SARIMA (0,2,2) (1,1,1) [12] model, which minimized RMSE, yielded the best results, with the 90/10 test set configuration showing the largest RMSE value, emphasizing the impact of training data size on model accuracy.

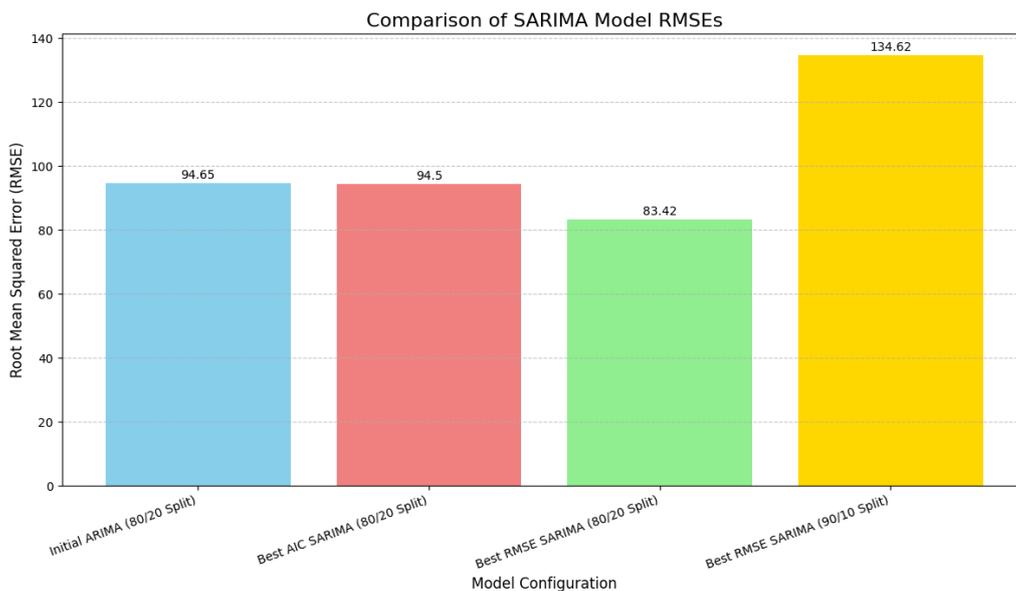


Figure 6. Comparison of SARIMA Model RMSEs.

4.2. Prophet Model Performance

The Facebook Prophet model was evaluated using the same test set that was used for the SARIMA model. Prophet's forecast was generated using its default settings, and the RMSE for this model was found to be 136.01, which is slightly worse than the SARIMA model's RMSE of 134.62. This suggests that while Prophet is a powerful tool, it did not perform as well as SARIMA in forecasting AI job market trends on this dataset. Figure 7 visually compares the training data, test data, and Prophet's forecast. The plot shows that Prophet was able to capture some of the trends but diverged significantly in later periods, particularly in the more recent forecast horizons.

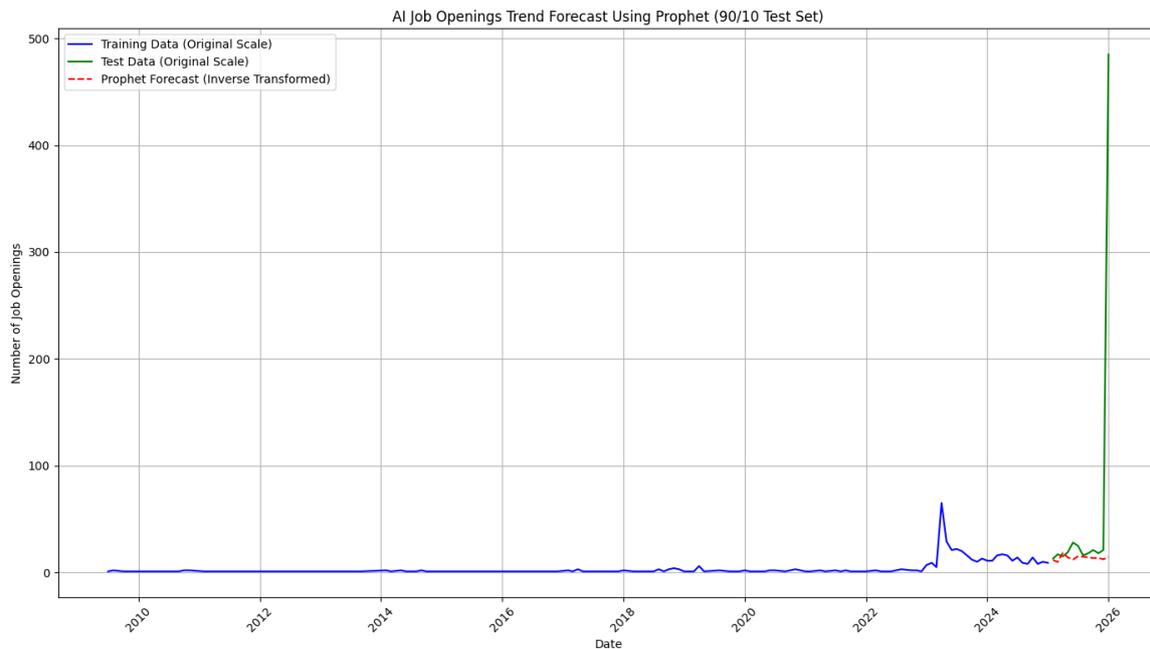


Figure 7. AI Job Openings Trend Forecast Using Prophet (90/10 Test Set)

Prophet's ability to model non-linear trends and seasonality is particularly useful when the data exhibits irregular fluctuations, but in this case, the SARIMA model with carefully tuned parameters outperformed it. Despite Prophet's strengths in handling missing data and outliers, the RMSE comparison suggests that for this particular time series dataset, SARIMA was more accurate in predicting the AI job openings trend.

4.3. Model Comparison

In comparing the SARIMA and Prophet models, it is evident that SARIMA slightly outperformed Prophet in terms of forecasting accuracy as measured by RMSE. Figure 8 highlights the RMSE values for both models on the 80/20 split and 90/10 split. The SARIMA model consistently showed lower RMSE values compared to Prophet, which indicates that, for this dataset, SARIMA's seasonality handling and parameter optimization provided a better fit for the data than Prophet's flexible modeling approach.

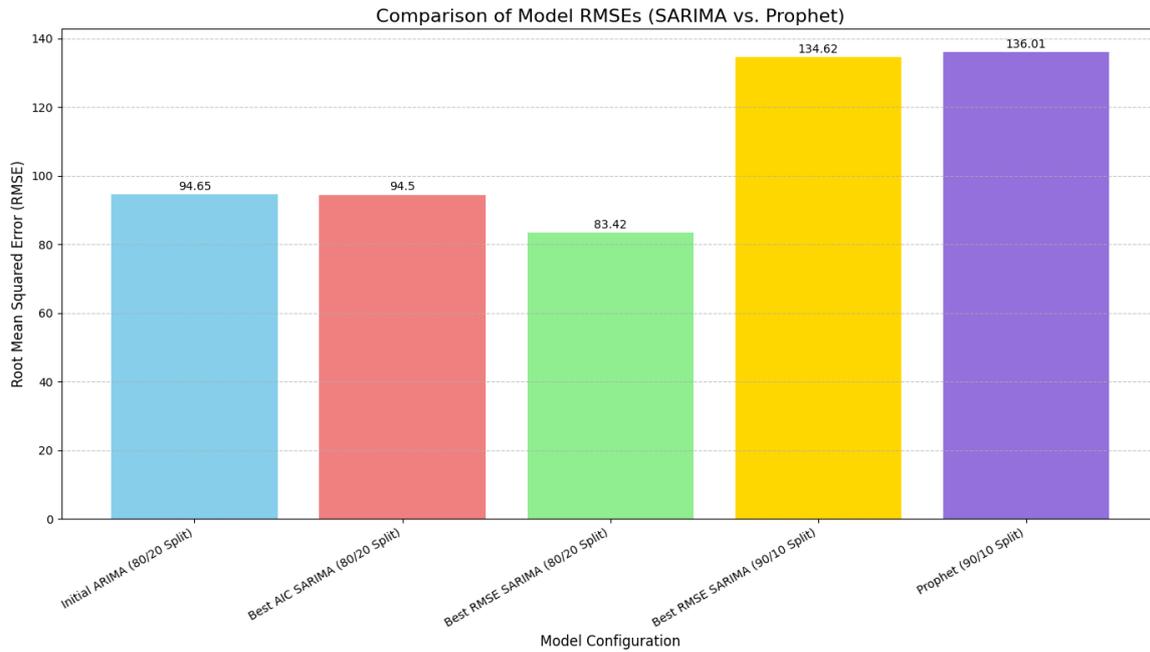


Figure 8. Comparison of Model RMSEs (SARIMA vs. Prophet)

The bar chart in [Figure 8](#) demonstrates the RMSE comparison across different configurations of SARIMA, including the initial ARIMA model, the SARIMA model with the lowest AIC, the SARIMA model with the lowest RMSE, and the final optimized SARIMA model on the 90/10 test set. This comparison illustrates that the optimized SARIMA configuration with an RMSE of 83.42 outperformed the Prophet model's RMSE of 136.01, reinforcing the conclusion that SARIMA was the more reliable model for this particular job market trend prediction task.

In summary, while Prophet provides a robust framework for handling missing data and non-linear trends, SARIMA's performance in this case shows the importance of careful parameter tuning and seasonality modeling in forecasting AI job openings. SARIMA's superior performance highlights the importance of understanding the data's characteristics and selecting the most suitable model accordingly.

4.4. Cross-Validation Results

To further evaluate the robustness of the SARIMA model, TSCV was implemented using a rolling forecast origin method. This approach mirrors real-world forecasting scenarios by training the model on a growing window of historical data and testing it on the subsequent data points. The mean RMSE across 14 folds was found to be 19.677, with a standard deviation of 32.273. This variability indicates that while there were some fluctuations in the model's performance across individual forecast horizons, the average performance remained consistently better than the single-split RMSE values.

[Figure 9](#) illustrates the RMSE values for each fold. The mean RMSE is represented by the red dashed line, and the dotted green lines show one standard deviation above and below the mean. As seen in the figure, the RMSE remains relatively stable for the first 12 folds, but there is a sharp spike in fold 14, which indicates an outlier or an unexpected deviation in the forecasted values for that period. Despite this, the overall trend of the RMSE across folds suggests that the model is generally robust and reliable for predicting AI job openings over time.

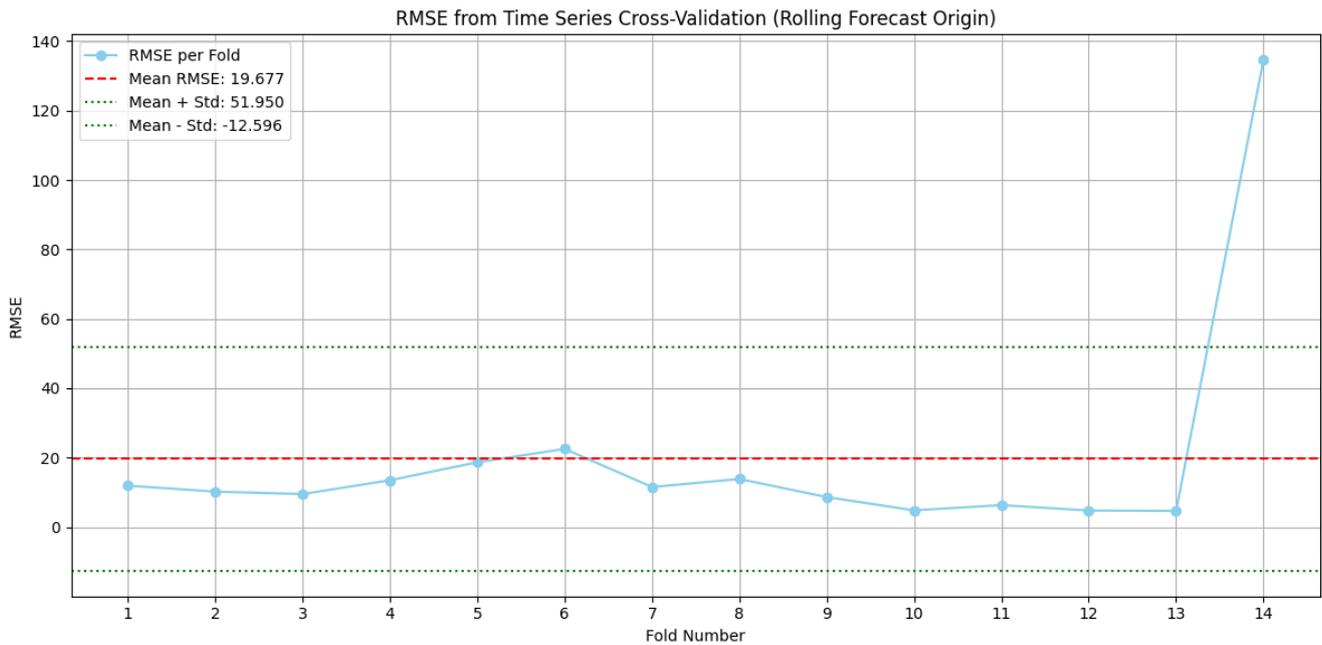


Figure 9. RMSE from Time Series Cross-Validation (Rolling Forecast Origin)

This cross-validation process provided a more robust estimate of the SARIMA model’s performance compared to the initial single-split evaluation. It highlighted that while individual forecast horizons may exhibit high variability, the model performed consistently within an acceptable range of error across multiple time periods. This further reinforces the SARIMA model’s capacity to generalize across different forecasting scenarios.

5. Discussion

5.1. Analysis of Model Performance

The comparison between the SARIMA and Prophet models reveals interesting insights into the forecasting of AI job market trends. As highlighted in the literature [15], [16], SARIMA’s strength lies in its ability to handle seasonal variations, making it a valuable tool when forecasting trends with clear periodic fluctuations. In this study, the SARIMA (0,2,2) (1,1,1) [12] model significantly outperformed Prophet, with the SARIMA model showing an RMSE of 83.42 compared to Prophet’s 136.01 on the 90/10 test set, as depicted in Figure 2. This confirms SARIMA’s advantage when seasonal patterns are strong and well-defined, in line with findings from previous studies that emphasize SARIMA’s suitability for seasonal time series data [18].

On the other hand, Prophet’s flexibility in handling non-linear trends and irregular data [19] is reflected in its performance, though it was not as effective as SARIMA in this specific dataset. Figure 5 shows the Prophet forecast for the AI job market, where the model captures some trends but diverges significantly in the later periods, particularly with respect to the forecast horizons. While Prophet is known for its ease of use and minimal parameter tuning [20], its inability to outperform SARIMA in this study indicates that for data with clear seasonality, a more structured model like SARIMA is preferable.

The results are consistent with the findings of Sultana [5] that SARIMA excels in handling seasonal data with clear periodicity, while Prophet shines in more complex datasets with non-linear trends and missing data. The comparison of RMSE values, shown in Figure 2, supports this conclusion, highlighting SARIMA’s superior performance when seasonality is a dominant factor.

5.2. Practical Implications

The results from this study have important practical implications for policymakers and business leaders who rely on accurate forecasting for AI job market trends. The SARIMA model’s superior performance indicates that it is well-suited for predicting job openings in sectors where seasonal trends and recurring patterns dominate. For example,

policymakers can use SARIMA to forecast job openings in sectors like data analysis and AI development, which exhibit clear seasonal fluctuations in demand. Understanding these trends can help in workforce planning, ensuring that training programs and educational resources align with anticipated job market needs [13].

However, businesses operating in more dynamic sectors or those experiencing significant disruptions may benefit from Prophet's flexibility. For instance, AI-driven companies or startups, where trends may shift quickly, could utilize Prophet to handle more unpredictable, non-linear data. Moreover, the adaptability of Prophet in handling missing data and outliers [19] could prove useful in industries facing sporadic data or sudden shifts in market conditions, particularly in the early stages of AI adoption.

In light of these findings, stakeholders should choose the model best suited to their specific forecasting needs. For industries with clear, well-understood seasonal patterns, SARIMA is the more reliable choice. Conversely, for industries with volatile and non-linear data, Prophet's ease of use and ability to handle irregular trends may provide better forecasting accuracy.

5.3. Limitations and Future Research

While the SARIMA model demonstrated robust performance, several limitations must be considered. One of the main challenges lies in the model's reliance on the availability of historical data, as the accuracy of SARIMA predictions depends heavily on past patterns. Additionally, while the SARIMA model is powerful for seasonal data, it can struggle when faced with highly non-linear or irregular data, which was not a primary concern in this study but could pose a challenge in other contexts. The inability of the SARIMAX model, incorporating the exogenous variable, to improve forecasting accuracy further highlights the need for careful selection of external variables that directly impact AI job trends.

Future research could explore the incorporation of more meaningful exogenous variables, such as economic indicators or technological trends, which may provide additional insight into the AI job market. Integrating these variables could help improve forecasting accuracy, especially in sectors influenced by external economic or policy factors. Another promising avenue for future research involves exploring machine learning models, particularly Long Short-Term Memory (LSTM) networks, which are designed to handle long-term dependencies and non-linear trends more effectively. This could provide further improvements in forecasting, particularly for AI job market trends in highly volatile environments.

Future studies could also incorporate a broader range of evaluation metrics, such as MAE and MAPE, to provide a more comprehensive assessment of model performance across different contexts [26]. By addressing these limitations and expanding the models tested, future research could refine AI job market forecasting, offering more reliable tools for stakeholders in policy, business, and workforce planning.

6. Conclusion

This study compared the SARIMA and Prophet models for forecasting AI job market trends, with SARIMA consistently outperforming Prophet in terms of RMSE, especially when seasonal patterns were present. The SARIMA (0,2,2) (1,1,1) [12] model achieved the best performance with an RMSE of 83.42, while Prophet's RMSE was 136.01. Cross-validation results confirmed SARIMA's robustness, with a mean RMSE of 19.677, highlighting its effectiveness for datasets influenced by seasonality.

For stakeholders like policymakers and workforce planners, the study suggests prioritizing SARIMA for sectors with clear seasonal trends, such as tech and AI-related fields. Prophet can be a good alternative for volatile industries due to its flexibility in handling non-linear trends and missing data. Integrating TSCV is also recommended for ensuring model robustness; while incorporating exogenous variables like economic or policy factors could further improve forecasting accuracy.

Time series forecasting is crucial for navigating the future of AI in the labor market. By leveraging models like SARIMA and Prophet, stakeholders can make more informed decisions in workforce planning, AI integration, and policy-making, ensuring the workforce is equipped for the challenges and opportunities posed by AI.

7. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization S.W. and B.S.; Methodology, S.W. and B.S.; Software, S.W.; Validation, S.W.; Formal Analysis, S.W.; Investigation, B.S.; Resources, S.W.; Data Curation, B.S.; Writing Original Draft Preparation, S.W.; Writing Review and Editing, S.W. and B.S.; Visualization, B.S. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] G. B. Isataeva, "Artificial Intelligence and the Transformation of Employment: A Systematic Literature Review," *Bull. Turan Univ.*, vol. 2025, no. 4, hal. 205–217, 2025, doi: 10.46914/1562-2959-2025-1-4-205-217.
- [2] A. Hassani dan M. Javadi, "Leveraging Artificial Intelligence to Study and Forecast Its Impacts on Employment Management and Wages in China's Key Cities," *Spectr. Eng. Manag. Sci.*, vol. 4, no. 1, hal. 43–54, 2025, doi: 10.31181/sems41202660h.
- [3] F. Mahmud *et al.*, "AI-Powered Workforce Analytics Forecasting Labor Market Trends and Skill Gaps for U.S. Economic Competitiveness," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 5, hal. 265–277, 2024, doi: 10.32996/jcsts.2024.6.5.21.
- [4] D. Liu, "Enterprise Digital Retail Business Data Analysis and Forecasting Based on Time Series Analysis," *Adv. Econ. Manag. Polit. Sci.*, vol. 77, no. 1, hal. 206–212, 2024, doi: 10.54254/2754-1169/77/20241678.
- [5] Fozia Sultana, Urooj Talpur, Muhammad Shahid Iqbal, Ahmed Ali, dan Khalid Hussain Memon, "The Macroeconomic Implications of Automation and AI on Labor Markets and Employment," *Crit. Rev. Soc. Sci. Stud.*, vol. 2, no. 2, hal. 497–507, Nov 2024, doi: 10.59075/f9hdkk61.
- [6] N. Masoud, "Artificial Intelligence and Unemployment Dynamics: An Econometric Analysis in High-Income Economies," *Technol. Sustain.*, vol. 4, no. 1, hal. 30–50, 2024, doi: 10.1108/techs-04-2024-0033.
- [7] W. Kao, "Impaction of Artificial Intelligence on the Labor Market," *Adv. Econ. Manag. Polit. Sci.*, vol. 84, no. 1, hal. 24–29, 2024, doi: 10.54254/2754-1169/84/20240772.
- [8] R. Jadhav dan A. Banubakode, "The Implications of Artificial Intelligence on the Employment Sector," *Int. J. Multidiscip. Res.*, vol. 6, no. 3, hal. 1–11, 2024, doi: 10.36948/ijfmr.2024.v06i03.22716.
- [9] R.-J. Yan, "The Impact of Artificial Intelligence on the Labor Market," *Int. J. Glob. Econ. Manag.*, vol. 2, no. 1, hal. 233–238, Jan 2024, doi: 10.62051/ijgem.v2n1.29.
- [10] R. Fu, "Analysis of the Effect of Artificial Intelligence on the Labor Market in the United States," *Adv. Econ.*

- Manag. Polit. Sci.*, vol. 57, no. 1, hal. 78–84, 2024, doi: 10.54254/2754-1169/57/20230538.
- [11] E. Dahlin, “Who Says Artificial Intelligence Is Stealing Our Jobs?,” *Socius Sociol. Res. a Dyn. World*, vol. 10, no. 1, hal. 1–13, 2024, doi: 10.1177/23780231241259672.
- [12] R. Shengelia, “Artificial Intelligence and Labor Market Dynamics: Employment Problems and Development Trends,” *Economics*, vol. 107, no. 3–5, hal. 7–13, 2025, doi: 10.36962/ecs107/3-5/2025-07.
- [13] P. D. Nalwade, “Adapting to Tomorrow’s Workforce: Navigating the Impacts of Artificial Intelligence on Employment,” *Interantional J. Sci. Res. Eng. Manag.*, vol. 08, no. 04, hal. 1–5, 2024, doi: 10.55041/ijrsrem30664.
- [14] A. Havryk dan T. Hazapova, “Artificial Intelligence and Its Role in the Labor Market and Financial Sector Itself: US Point of View,” *Int. Sci. J. Manag. Econ. Financ.*, vol. 3, no. 3, hal. 1–9, 2024, doi: 10.46299/j.isjmef.20240303.01.
- [15] Y. Zhao, “E-Commerce Demand Forecasting Using SARIMA Model and K-Means Clustering Analysis,” *J. Innov. Dev.*, vol. 7, no. 1, hal. 1–6, 2024, doi: 10.54097/ctfb0379.
- [16] V. Arumugam dan V. Natarajan, “Time Series Modeling and Forecasting Using Autoregressive Integrated Moving Average and Seasonal Autoregressive Integrated Moving Average Models,” *Instrum. Mes. Métrologie*, vol. 22, no. 4, hal. 161–168, 2023, doi: 10.18280/i2m.220404.
- [17] M. Othman, R. Indawati, A. A. Suleiman, M. B. Qomaruddin, dan R. Sokkalingam, “Model Forecasting Development for Dengue Fever Incidence in Surabaya City Using Time Series Analysis,” *Processes*, vol. 10, no. 11, hal. 2454, 2022, doi: 10.3390/pr10112454.
- [18] G. Dudek, “STD: A Seasonal-Trend-Dispersion Decomposition of Time Series,” *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 10, hal. 10339–10350, Okt 2023, doi: 10.1109/TKDE.2023.3268125.
- [19] F. L. Aires, G. D. Galeno, F. N. Belchior, A. M. Oliveira, dan J. D. Hunt, “Enhancing three-phase induction motor reliability with health index and artificial intelligence-driven predictive maintenance,” *R. Soc. Open Sci.*, vol. 12, no. 5, hal. 1–21, Mei 2025, doi: 10.1098/rsos.241946.
- [20] K. Albeladi, B. Zafar, dan A. Mueen, “A novel deep-learning based approach for time series forecasting using sarima, neural prophet and fb prophet,” *Rev. Gestão Tecnol.*, vol. 24, no. Special, hal. 72–86, Mei 2024, doi: 10.20397/2177-6652/2024.v24iSpecial.2818.
- [21] G. V. S. A. Varma, G. V. N. A. Varma, dan A. Kethaan, “Predictive Insights for Monthly Property Sales Forecasting: An End-to-End Time Series Forecasting,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 1, hal. 598–604, 2023, doi: 10.48175/ijarsct-12493.
- [22] S. Ryu, H. J. Nam, M. Jhon, J. Lee, J. Kim, dan S. Kim, “Trends in Suicide Deaths Before and After the COVID-19 Outbreak in Korea,” *PLoS One*, vol. 17, no. 9, hal. e0273637, 2022, doi: 10.1371/journal.pone.0273637.
- [23] Y. Shebalkova, J. K. Jacob, dan Y. Apanovych, “Silver and Palladium Price Dynamics: Analyzing Relationships and Forecasting with ARIMA and Prophet,” *Acta Montan. Slovaca*, vol. 30, no. 1, hal. 170–184, Jul 2025, doi: 10.46544/AMS.v30i1.13.
- [24] K. Karthick, S. Aruna, R. Shanmugasundaram, M. Safran, S. Alfarhood, dan W. Al-Rajhi, “Enhancing Sustainable Urban Energy Management Through Short-Term Wind Power Forecasting Using LSTM Neural Network,” *Sustainability*, vol. 15, no. 18, hal. 13424, 2023, doi: 10.3390/su151813424.
- [25] J. Tan, “NVIDIA Stock Price Prediction by Machine Learning,” *Highlights Business, Econ. Manag.*, vol. 24, no. 1, hal. 1072–1076, Jan 2024, doi: 10.54097/dsz8ns50.
- [26] N. S. Wen dan L. S. Ling, “Evaluation of Cryptocurrency Price Prediction Using LSTM and CNNs Models,” *JOIV Int. J. Informatics Vis.*, vol. 7, no. 3–2, hal. 2016, Nov 2023, doi: 10.30630/joiv.7.3-2.2344.
- [27] S. Kwarteng dan P. Andreevich, “Comparative Analysis of ARIMA, SARIMA and Prophet Model in Forecasting,” *Res. Dev.*, vol. 5, no. 4, hal. 110–120, Okt 2024, doi: 10.11648/j.rd.20240504.13.

- [28] E. Spiliotis, F. Petropoulos, dan V. Assimakopoulos, “On the Disagreement of Forecasting Model Selection Criteria,” *Forecasting*, vol. 5, no. 2, hal. 487–498, 2023, doi: 10.3390/forecast5020027.
- [29] T. S. Talagala, R. J. Hyndman, dan G. Athanasopoulos, “Meta-learning How to Forecast Time Series,” *J. Forecast.*, vol. 42, no. 6, hal. 1476–1501, 2023, doi: 10.1002/for.2963.
- [30] F. Petropoulos dan E. Siemsen, “Forecast Selection and Representativeness,” *Manage. Sci.*, vol. 69, no. 5, hal. 2672–2690, 2023, doi: 10.1287/mnsc.2022.4485.
- [31] A. Jakovljevic, L. Charlin, dan B. Barbeau, “Applying Recurrent Neural Networks and Blocked Cross-Validation to Model Conventional Drinking Water Treatment Processes,” *Water*, vol. 16, no. 7, hal. 1042, Apr 2024, doi: 10.3390/w16071042.