

Machine Learning Approaches for Temperature Forecasting in Sialkot, Pakistan: A Comparative Study with Traditional Methods

Wajiha Nasir (Corresponding Author)

Department of Statistics, GC Women University, Sialkot, Lahore, Pakistan

Email: wajiha.nasir@gcwus.edu.pk

Mozama Jaffar

Department of Statistics, GC Women University, Sialkot, Lahore, Pakistan

Email: mozamajaffar@gmail.com

Alina Ramzan

Department of Statistics, GC Women University, Sialkot, Lahore, Pakistan

Email: wajiha.nasir@gcwus.edu.pk

Ayesha Khalid

Department of Statistics, GC Women University, Sialkot, Lahore, Pakistan

Email: aishaa.kh89@gmail.com

ABSTRACT

This study presents a comparative analysis of traditional time series and machine learning (ML) models for short-term temperature forecasting in Sialkot, Pakistan. Using a decade-long dataset (2015–2025) that includes daily temperature, humidity, and atmospheric pressure, univariate and multivariate models were developed. Classical methods such as Simple Exponential Smoothing and Holt-Winters yielded high error rates (MAPE = 27%). In contrast, the Artificial Neural Network (ANN) model achieved the best precision (MAPE = 2.97%, RMSE = 2.83, R² = 0.960), closely followed by LSTM and Linear Regression. The results highlight the strength of ML for localized weather forecasting, with implications for agriculture, urban planning, and the reduction of disaster risk.

Keywords: Temperature Forecasting, Machine Learning, Time Series Models, Artificial Neural Network,

Introduction

Weather forecasting plays a crucial role in various sectors, including agriculture, disaster management, infrastructure planning, and public health. Accurate predictions of key meteorological variables, such as temperature, humidity, and atmospheric pressure, are essential to mitigate the adverse impacts of extreme weather events and ensure efficient resource allocation. Traditionally, Numerical Weather Prediction (NWP) models have been used for this purpose. Although these models are robust and physics-based, they are computationally intensive and often struggle to capture

localized climatic variations, especially in rapidly changing environments. Sialkot, an industrial and agricultural hub in northeastern Pakistan, experiences a subtropical climate marked by hot summers, monsoon rains, and mild winters. The region's socioeconomic activities, particularly agriculture and manufacturing, are highly sensitive to weather variability. Therefore, accurate and timely weather forecasts are critical for informed decision-making and economic resilience in this region. In recent years, machine learning (ML) has emerged as a powerful alternative to traditional statistical and numerical forecasting methods. ML algorithms can model complex, nonlinear relationships in historical data and often deliver higher predictive accuracy with reduced computational requirements. Models such as Random Forest, Gradient Boosting, Support Vector Regression (SVR), Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks have shown promising results in various meteorological forecasting tasks worldwide. In this paper, we are comparing traditional time series models with some machine learning models. Early efforts in environmental modeling—such as air quality and ozone concentration forecasting—relied heavily on classical statistical methods. Comparative analyses of these approaches revealed limitations in capturing complex pollutant trends. Recent studies have extended these insights to temperature forecasting, a task vital for managing agriculture, energy consumption, urban infrastructure, and disaster preparedness. Abdullah et al. (2020) forecasted rainfall in Indonesia using both statistical methods (e.g., SARIMA) and ML models, evaluating their performance using Mean Absolute Percentage Error (MAPE). Saa and Ranathunga (2020) predicted the temperature in Hungary using ARIMA and deep learning models, finding the latter more precise than traditional approaches. Matta et al. (2021) compared multilayer perceptron (MLP), radial basis function (RBF) neural networks, and Gaussian process regression for economic forecasting using real-world data. Haque et al. (2021) applied deep neural networks to predict hourly temperatures in Bangladesh, reporting that GRU-LSTM achieved the lowest Root Mean Squared Error (RMSE), while Convolutional Neural Networks (CNN) offered computational efficiency. Their findings also highlighted the influence of dataset variability—such as variance and mean absolute deviation—on model performance across different regions. Gong et al. (2022) demonstrated that deep learning models like ConvLSTM and Stochastic Adversarial Video Prediction (SAVP) are effective for short-term temperature forecasting. SAVP, in particular, outperformed ConvLSTM in both accuracy and spatial pattern prediction due to its advanced generative architecture and use of auxiliary predictors like 850 hPa temperature. Although these models have not yet surpassed NWP in all respects, they hold strong potential for data-driven forecasting. Other researchers have applied similar techniques in different domains. Akyildirim et al. (2022) compared various models for forecasting stock prices based on historical data. Similarly, Sirisha et al. (2022) used ARIMA, SARIMA, and deep learning models to forecast profits, evaluating performance through multiple error metrics. Despite the growing success of ML techniques in weather prediction, comprehensive studies comparing machine learning algorithms with classical time series models for temperature forecasting in Sialkot are scarce. Moreover, the impact of incorporating multivariate features—such as humidity, pressure, and rainfall—on predictive accuracy remains underexplored for this region.

- Developing univariate and multivariate forecasting models for daily temperature prediction in Sialkot using historical weather data.
- Comparing the predictive performance of traditional time series models (SES and Holt-Winters) with advanced ML algorithms (ANN, LSTM, Random Forest, SVR, XGBoost, etc.).
- Identifying the most accurate and computationally efficient model for short-term temperature forecasting.

Methodology

This section outlines the study area, the characteristics of the dataset, the pre-processing steps, the development of the model, and the evaluation metrics used to forecast the temperature in Sialkot, Pakistan.

Study Area

Sialkot is located in the northeast of Punjab, Pakistan (32.5 ° N, 74.5 ° E). The city experiences a humid subtropical climate with different seasonal variations: hot summers, monsoon rains (July-September) and mild winters. These climatic patterns significantly influence agricultural and industrial activities, necessitating accurate weather predictions for local decision-making.

Data Collection

Historical meteorological data were collected from Weather Underground, spanning January 2015 to January 2025. The dataset includes:

- Temperature (°C) – target variable.
- Relative humidity (%) - explanatory variable.
- Atmospheric Pressure (inHg) – explanatory variable.

Data was recorded at daily intervals, ensuring a high-resolution time series suitable for short-term forecasting.

Forecasting Model

Two categories of forecasting models were implemented:

a) Classic Time Series Models

- Simple Exponential Smoothing (SES): Univariate method emphasizing recent observations.
- Holt-Winters Method: Incorporates trend and seasonality.

b) Machine Learning Models

- Linear Regression (LR): Baseline linear model.
- Random Forest (RF): Ensemble based regression approach.
- Support Vector Regression (SVR): Kernel-based method for capturing nonlinear patterns.
- XGBoost: Gradient boosting model for enhanced accuracy.
- K-Nearest Neighbors (KNN): Instance-based learning algorithm.
- Decision Tree (DT): Tree-based regression.
- Artificial Neural Network (ANN): Feed-forward network optimized using backpropagation.
- Long Short-Term Memory (LSTM): Deep learning model designed for sequential data.

80% of the dataset was used for training, and 20% for testing. For machine learning models, 5-fold cross-validation was applied. Grid search and random search were used for models such as XGBoost, SVR, and ANN. Models were

implemented in Python using Scikit-learn, TensorFlow, and Statsmodels libraries.

Evaluation Metrics

The performance of all models was evaluated using the following metrics:

- Mean Absolute Percentage Error (MAPE): Penalizes larger errors more heavily, indicating the magnitude of prediction error.
- Root Mean Squared Error (RMSE): Provides a straightforward measure of average absolute deviation.
- Mean Absolute Error (MAE): Provides a straightforward measure of average absolute deviation.
- Mean Squared Error (MSE): Quantifies the average of squared differences between actual and predicted values.
- Coefficient of Determination (R^2): Indicates the proportion of variance in the dependent variable explained by the model.

This comprehensive methodology enables the development and comparative evaluation of both traditional and modern approaches to temperature forecasting, providing actionable insights for climate-responsive urban management.

Results and Discussion

Exploratory Data Analysis

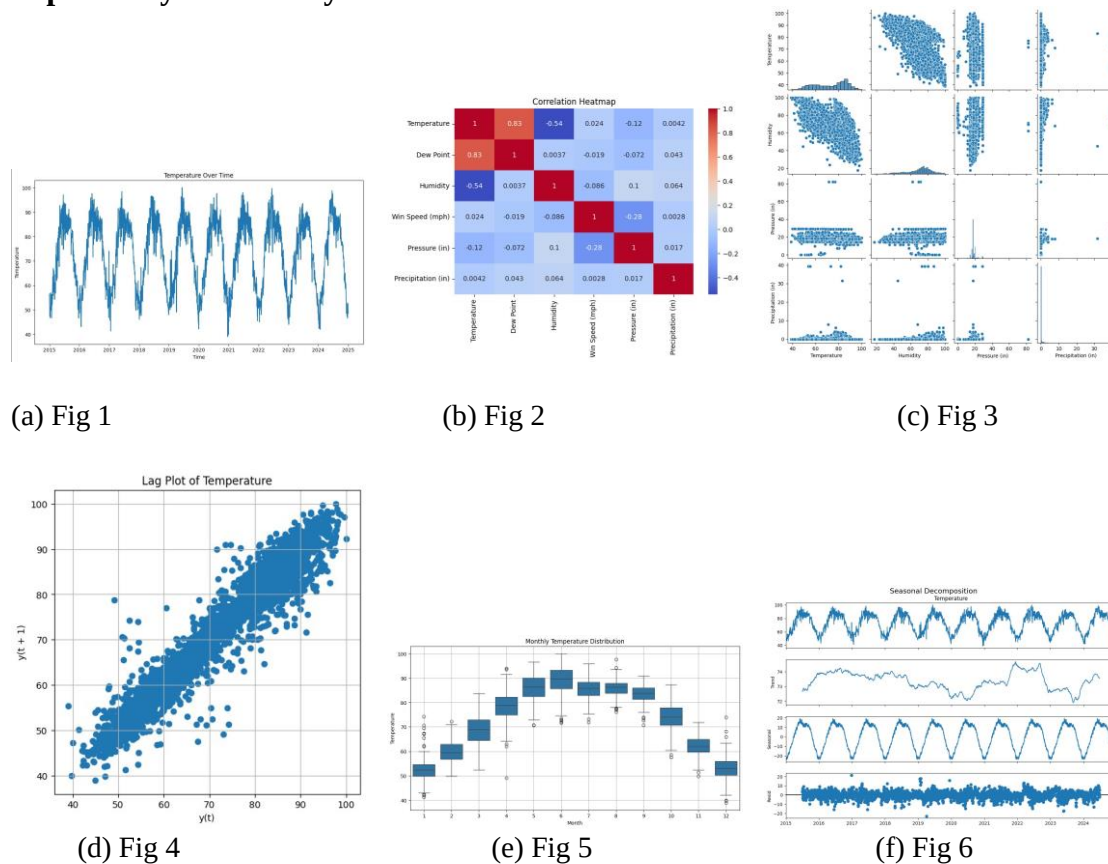


Figure 1: Grid of Forecasting Result Figures

Figure 1 provides an overview of the temperature time series and associated variables. Panel (a) illustrates the raw temperature series, revealing pronounced

seasonal fluctuations and long-term trends indicative of non-stationarity. Such characteristics confirm the necessity of incorporating seasonality in forecasting models. Panel (b) presents the correlation matrix, where temperature exhibits a strong positive correlation with relative humidity and moderate negative correlation with atmospheric pressure. These associations justify the inclusion of multiple predictors in multivariate models. The scatter plot matrix in Panel (c) highlights linear and non-linear relationships among variables, while the lag plot in Panel (d) confirms strong autocorrelation in temperature, supporting the adoption of autoregressive structures. Panel (e) and (f) further emphasize seasonal components, as demonstrated by boxplot distributions and trend-seasonal decomposition.

Model Performance Comparison

Table 1: Comparison of Forecasting Models on Temperature Prediction

Model	MAPE (%)	RMSE	MAE	MSE	R-Squared
SES	27.921	25.731	22.254	662.080	0.380
Holt-Winter	27.899	25.713	22.237	661.177	0.493
ANN	2.978	2.837	2.128	2.128	0.960
LSTM	3.042	2.849	2.132	8.117	0.956
Linear Regression	2.996	2.801	2.133	7.847	0.960
Random Forest	3.164	2.918	2.231	8.514	0.956
Support Vector Regression	3.195	2.975	2.225	8.850	0.954
XGBoost	3.397	3.156	2.386	9.958	0.949
K-Nearest Neighbors	3.493	3.204	2.479	10.267	0.947
Decision Tree	4.504	4.184	3.204	17.508	0.910

Table 1 summarizes the predictive performance of traditional time series methods and machine learning algorithms across multiple evaluation metrics (MAPE, RMSE, MAE, MSE, and R^2).

Discussion of Results

The results indicate that machine learning models significantly outperformed traditional time series methods in all metrics. The Simple Exponential Smoothing (SES) and HoltWinters models recorded the highest errors, with MAPE values exceeding 27% and R^2 values below 0.50, underscoring their inability to capture complex, non-linear dependencies and multivariate relationships.

Conversely, the Artificial Neural Network (ANN) demonstrated the best overall performance, achieving the lowest MAPE (2.978%) and RMSE (2.837) along with the highest R^2 (0.960). The strong performance of ANN can be attributed to its capacity to model intricate nonlinear patterns and interactions among predictors. Similarly, LSTM achieved competitive results (MAPE = 3.042%), benefiting from its ability to capture temporal dependencies in sequential data. Interestingly, Linear

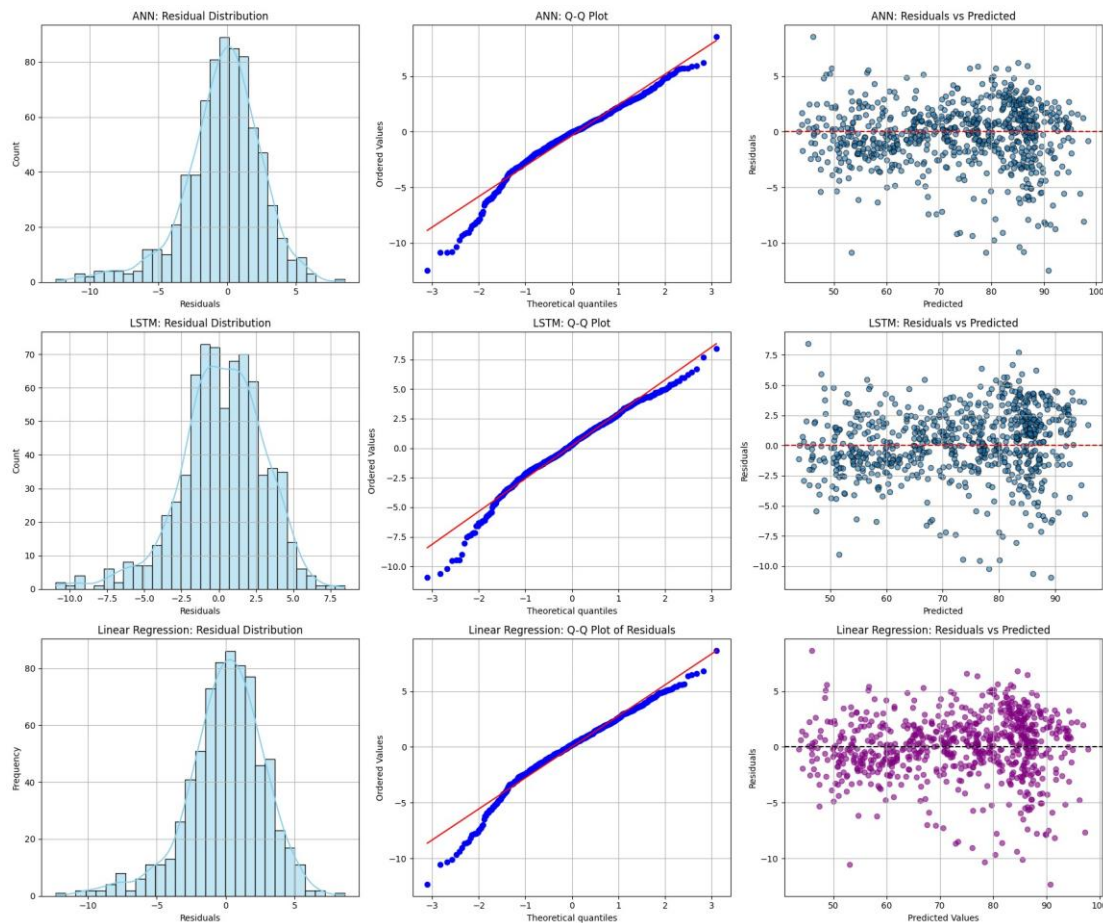


Figure 2: Residual Plots for ANN, LR and LSTM

Regression achieved comparable accuracy to ANN ($R^2 = 0.960$), indicating that the multivariate feature set exhibited a strong linear relationship with temperature. Ensemble methods such as Random Forest and XGBoost also delivered high accuracy, albeit slightly inferior to ANN and Linear Regression, likely due to their sensitivity to feature scaling and parameter tuning. In contrast, Decision Tree exhibited the weakest performance among ML models, with the highest error metrics and the lowest R^2 (0.910), suggesting susceptibility to overfitting and poor generalization.

Figure 2 presents a comparative residual analysis and actual-versus-predicted plots for the Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Linear Regression (LR) models, employed for temperature forecasting. The residual histograms for all three models approximate a normal distribution, centered around zero. Notably, the residuals from the ANN and LSTM models exhibit tighter clustering around the mean, indicating lower variance and a better fit to the observed data. In contrast, the Linear Regression model displays slightly more dispersion in its residuals, suggesting relatively higher prediction errors. The Quantile-Quantile (Q-Q) plots further assess the normality of residuals. Both the ANN and LSTM models exhibit points that align closely with the 45-degree reference line, confirming the

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

assumption of residual normality. The linear regression model, however, shows mild deviations at the tails, implying slight departures from normality, particularly for extreme values. Plots of residuals against predicted values for the ANN and LSTM models reveal no discernible patterns or trends, supporting the assumption of homoscedasticity and suggesting that the models adequately capture the underlying structure of the data. Conversely, the residuals for the linear regression model display a modest funnel shape, indicative of heteroscedasticity and potential unmodeled non-linearities. The lower panel of Figure 2 illustrates the comparison between actual and predicted temperature values: The ANN model demonstrates a strong overlap between actual and predicted values, effectively capturing fluctuations and underlying patterns in the data. The LSTM model similarly provides a high-fidelity fit, accurately modeling temporal dynamics and periodic trends. The Linear Regression model, while capturing the general trend, shows more substantial deviations from the actual values, particularly during peak variations, reflecting its limitations in modeling complex, nonlinear relationships.

Conclusion

This study conducted a comprehensive comparative analysis of traditional statistical models and advanced machine learning algorithms for short-term temperature forecasting in Sialkot, Pakistan. Using a decade-long dataset incorporating daily records of temperature, relative

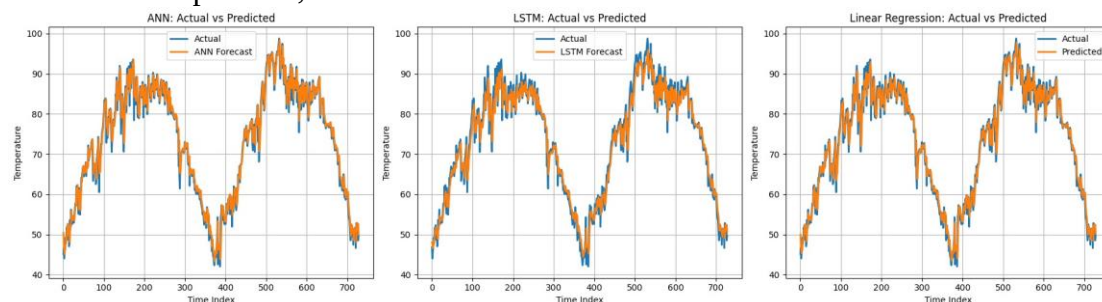


Figure 3: Actual vs Predicted Graph for ANN, LR and LSTM

humidity, and atmospheric pressure, both univariate and multivariate models were developed and evaluated using multiple error and performance metrics. The findings demonstrate that machine learning approaches significantly outperform classical time series models. The Artificial Neural Network (ANN) model achieved the highest predictive precision, with the lowest error rates (MAPE = 2.97%, RMSE = 2.83) and the highest value of R2 (0.960). The long-short-term memory (LSTM) and multivariate Linear Regression models also performed competitively, underscoring the strength of both deep learning and linear modeling under appropriate feature conditions. Traditional methods such as Simple Exponential Smoothing (SES) and Holt-Winters were found to be inadequate for capturing the nonlinear and multivariate dynamics of temperature data, reflected in their high error values and low explanatory power. In addition to model comparisons, the study also highlights the value of incorporating multivariate features, specifically humidity and pressure, which significantly improved forecast performance in all ML models. Residual analysis further validated the suitability of ANN and LSTM models, with well-

behaved error distributions and minimal bias. The results of this study have practical implications for agriculture, urban planning, disaster preparedness, and climate-sensitive policy making in regional contexts like Sialkot. By identifying computationally efficient and highly accurate models, this research supports the integration of machine learning tools into local weather forecasting systems. Future research may focus on: Extending the forecasting horizon to medium- and long-term temperature predictions. Incorporating additional meteorological and environmental variables such as rainfall, wind speed, and solar radiation. Exploring ensemble learning or hybrid models that combine the strengths of multiple algorithms. Evaluating real-time deployment and automated updating systems for operational forecasting. Ultimately, this work contributes to the growing body of literature advocating for data-driven, regionally adapted forecasting solutions that are both scalable and interpretable for decision-makers in developing regions. While the ANN and LSTM models performed exceptionally well, their interpretability and computational requirements may limit deployment in low-resource environments. Future studies may explore interpretable ML models or hybrid approaches combining domain knowledge with data-driven forecasting.

References

- AS Abdullah, BN Ruchjana, IG NM Jaya, et al. Comparison of sarima and svm model for rainfall forecasting in bogor city, indonesia. In *Journal of Physics: Conference Series*, volume 1722, page 012061. IOP Publishing, 2021.
- Erdinc Akyildirim, Aurelio F Bariviera, Duc Khuong Nguyen, and Ahmet Sensoy. Forecasting high-frequency stock returns: a comparison of alternative methods. *Annals of Operations Research*, 313(2):639–690, 2022.
- N. Asgari, M. Mahdavi, and S. Al-Ameri. Air pollution prediction using big data and machine learning. In *2017 3rd International Conference on Web Research (ICWR)*, pages 165–171. IEEE, 2017.
- D. Y. Chang, J. C. Chen, and S. J. Lee. A big data system for air quality analysis and prediction. *IEEE Access*, 6:21993–22002, 2018.
- X. Chen, X. Li, S. Ma, and X. Tang. Big data, information theory, and machine learning for air quality research in developing countries. *Environmental Science & Technology*, 52(1):154–162, 2018.
- Eranga De Saa and Lochandaka Ranathunga. Comparison between arima and deep learning models for temperature forecasting. *arXiv preprint arXiv:2011.04452*, 2020.
- C. Esposito, J. Lim, and C. Li. Improving the accuracy of low-cost air quality sensors using neural networks. In *IEEE Sensors Journal*, volume 16, pages 6934–6941, 2016.
- E. M. Ghoneim, N. El-Sheimy, and S. El-Rabaie. Deep learning for ozone concentration prediction. *Environmental Pollution*, 228:172–181, 2017.
- Bing Gong, Michael Langguth, Yan Ji, Amirpasha Mozaffari, Scarlet Stadler, Karim Mache, and Martin G Schultz. Temperature forecasting by deep learning methods. *Geoscientific model development*, 15(23):8931–8956, 2022.
- Ehtashamul Haque, Sanzana Tabassum, and Eklas Hossain. A comparative analysis of deep neural networks for hourly temperature forecasting. *IEEE Access*, 9:160646–160660, 2021.
- M. A. Ishak, N. H. Omar, and N. M. Nasir. Comparative analysis of statistical methods for ozone concentration forecasting. In *2017 IEEE Conference on Energy*

Online ISSN

3007-3197

Print ISSN

3007-3189

<http://amresearchreview.com/index.php/Journal/about>

Conversion (CENCON), pages 37–41. IEEE, 2017.

M. Keller, Y. Li, and Q. Li. Machine learning for air quality modeling: A review. *Atmospheric Environment*, 153:1–13, 2017.

T. Li, Y. Zhang, and C. Li. Deep learning for air quality prediction. *Journal of Cleaner Production*, 112:1836–1844, 2016.

Cláudia Eliane da Matta, Natália Maria Puggina Bianchesi, Milena Silva de Oliveira, Pedro Paulo Balestrassi, and Fabiano Leal. A comparative study of forecasting methods using real-life econometric series data. *Production*, 31:e20210043, 2021.

P. Nowack, J. Messerschmidt, S. Ma, and X. Tang. A machine learning approach to quantify meteorological drivers of ozone pollution in china from 2015 to 2018. *Atmospheric Environment*, 269:118835, 2022.

Z. Peng, C. Zhang, and J. Li. Hourly air quality prediction using machine learning techniques. *Atmospheric Environment*, 162:106–116, 2017.

Y. Rybarczyk, A. Czerwin´ska, and R. Czerwin´ski. Machine learning approaches for short-term air pollution forecasting. *Environmental Science and Pollution Research*, 31(1):143–158, 2024.

A. Shamsuddin and S. Shahid. High-resolution pm2.5 prediction using machine learning algorithms: A review. *Environmental Science and Pollution Research*, 30(1):1–18, 2023.

Uppala Meena Sirisha, Manjula C Belavagi, and Girija Attigeri. Profit prediction using arima, sarima and lstm models in time series forecasting: A comparison. *Ieee Access*, 10:124715–124727, 2022.

World Bank. The cost of air pollution: Strengthening the economic case for action. World Bank Publications, 2020.

R. Yan, X. Li, and C. Li. A novel neural network model for air pollution forecasting. *IEEE Transactions on Neural Networks and Learning Systems*, 28(6):1361–1372, 2017.

Y. Zhang and X. Yuan. Big data analytics for air quality prediction. In *2015 IEEE International Conference on Big Data (Big Data)*, pages 190–195. IEEE, 2015.