Research Article/Araştırma Makalesi

GAZİ
JOURNAL OF ENGINEERING SCIENCES

Time Series Prediction of Temperature Using Seasonal ARIMA and LSTM Models

Hakan Koçak[*a]

Submitted: 18.05.2023 Revised: 25.08.2023 Accepted: 29.09.2023 doi:10.30855/gmbd.0705088

Keywords: Time series prediction, monthly mean temperature, ARIMA, SARIMA, LSTM, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF)

[a] Meteoroloji Genel Müdürlüğü – 06120 Ankara, Türkiye
Orcid: 0000-0003-2491-327X
e mail: hkocak@gmail.com

[*] Corresponding author: hkocak@gmail.com

ABSTRACT

Precise quantitative understanding and monitoring of temperature is indispensable due to its tremendous impact on almost every aspect of our lives. This work investigates prediction capabilities of two machine learning techniques, namely Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) and compares them in predicting monthly mean temperature time series data for a weather station in Ankara, Türkiye from January 2010 to March 2023. The comparison of forecasting performance was based on mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE). The results showed that both models can capture the variations of time series data. Both models exhibited reasonably good performance in predicting monthly mean temperature, but the ARIMA model gave the least forecast error compared to the LSTM model.

ARIMA ve LSTM Modelleri ile Sıcaklık Zaman Serileri Tahmini

ÖZ

Hava sıcaklığının insan hayatının hemen her alanındaki büyük öneminden dolayı sıcaklığın nicel olarak anlaşılması ve takip edilmesi oldukça elzemdir. Bu çalışmada Ankara'da bulunan bir meteorolojik hava istasyonundan elde edilen Ocak 2010 ila Mart 2023 tarihleri arasındaki gözlem verileri kullanarak Mevsimsel Otoregresif Entegre Hareketli Ortalama (Seasonal Autoregressive Integrated Moving Average -SARIMA) ve Uzun Kısa-Vadeli Hafıza Ağları (Long Short-Term Memory-LSTM) makine öğrenmesi metotlarıyla aylık ortalama hava sıcaklığının kestirimi yapılmış ve bu metotların sıcaklık tahmini konusundaki başarısı karşılaştırılarak irdelenmiştir. Modellerin tahmin başarıları Ortalama Karesel Hata (OKH), Karekök Ortalama Karesel Hata (KOKH) ve Ortalama Mutlak Hata (OMH) performans metrikleri kullanarak yapılmıştır. Araştırma sonucunda her iki yöntemin de aylık ortalama sıcaklık kestiriminde iyi derecede performans gösterdiği ortaya çıkmıştır. Bununla birlikte ARIMA modelinin LSTM modeline göre hata oranının kriter olarak ele alınan üç metrik (OKH, KOKH, OMH) için de daha düşük olduğu, dolayısıyla daha iyi performansa sahip olduğu görülmüştür.

Anahtar Kelimeler: Zaman serileri tahmini, aylık ortalama sıcaklık, ARIMA, SARIMA, LSTM, Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF)

To cite this article: H. Koçak, "Time Series Prediction of Temperature Using ARIMA and LSTM Models," *Gazi Journal of Engineering Sciences,* vol. 9, no. 3, pp. 574-584, 2023. doi:10.30855/gmbd.0705088

1. Introduction

Temperature is one of the important climatological parameters as it has direct as well as indirect effect on humans, animals, and plants. Due to climate change which is a result of greenhouse gas emissions due to human activities on earth, we see long-term shifts in temperatures and weather patterns. Asa a result of climate change, we see more severe and frequent weather events such as storms, flash floods, as well as increased drought, warming oceans, rising sea levels, food shortages, poverty and displacement, and food shortages.

There are many areas where ambient temperature is a key factor including agriculture, energy, water resources management, solar power. Deviations in the temperatures result in more energy demand which can lead to severe stress on the power grids. There is direct effect ambient temperature on growth of crops, crop yield, and water demand for growing crops in agriculture. Temperatures not suitable for the growth of crops can lead to poor yield as well as loss in productivity. Therefore, the importance of temperature remains as an important topic for researchers to conduct studies on it.

This study aims at predicting monthly temperature values based on historical temperature data for a weather station located in the city of Ankara, Türkiye from January 2010 to March 2023 using two time series prediction methodologies, namely Seasonal Integral Moving Average (SARIMA) as well as Short Long-Term Memory (LSTM) and comparing their prediction performances. ARIMA (and its extension SARIMA, which accounts for seasonality in the data), and LSTM models have gained an increasing popularity among researchers for time series analysis in many different areas (For example, see [1, 2, 3, 4, 5]). Thus, these two methodologies were selected for this study in order to assess and com-pare performances of the popular time-series analysis and forecasting methodologies.

Time series forecasting methods use historical data, analyse it to find recurrent patterns existing in the data. Data that has certain periodical characteristics in it, then used to forecast future events, and occurrences. Autoregressive Integrated Moving Average (ARIMA), which are of stochastic in nature, have gained quite a popularity as a time series forecasting model. An ARIMA model can be thought of as a different type of regression analysis model which estimates the relationships between a dependent variable and independent variables. LSTM models provide another approach to time series forecasting, which are widely used in various areas. LSTM networks are a variety of recurrent neural networks (RNSs) and used in the field of deep learning. One of the major drawbacks of the RNN networks is that they have short-term memory to retain previous information in the current neuron. To fix this, LSTM models were developed in order to be able to keep information for a longer period of time [1], [4], [6], [7], [8].

Both ARIMA and LSTM methods were applied successfully in a number of studies by various re-searchers to predict meteorological parameters including daily and monthly temperature values. [9] observed mean temperature changes at regional scale in Turkey from 1950 to 1994 using time-series analysis. They found a statistically colling trend at 21 stations, a warming trend at one station and no trend at 36 stations. [10] applied ARIMA model to data of four weather stations for the period of 1990 to 2011 in Iraq. [11] analysed monthly mean temperature in Nanjing, China. They used monthly mean temperature from 1951 to 2014 as the training set and data between 2015-2017 as the testing set and developed an ARIMA model for their study.

[7] used both Recurrent Neural networks (RNN) and LSTM methodologies to predict daily temperature and classify them into five categories namely "Cold", "Cool", "Normal", "Warm" and "Hot" using data from 2000 to 2019. [1] conducted a study of forecasting and modelling daily temperature for 4 European sites of different climatic zones based on data from 1980 to 2010. They used SARIMA, the autoregressive integrated moving average (ARIMA) with external regressors and showed the developed model was able to represent the data series and could be used for forecasting of future daily temperature values.

[12] employed a SARIMA model and forecasted mean temperature for the city of Gujarat, India using the past data from period of 1984 to 2015. They tried several models and selected the best model of SARIMA (1, 0, 1) (1, 1, 1) (12) for forecasting temperature values based on the Akaike Information Criterion (ACI). They tested the adequacy of their model and the model diagnostics showed that the model was reliable for forecasting monthly mean temperatures. [13] forecasted daily maximum temperature using three different methods, namely ARIMA, SARIMA and Autoregressive Fractional Integrated Moving Average (ARFIMA) for four different regions in Kerala, India based on data from January 2019 to December 2020. They then

compared the performances of three methods by using mean squared error (MSE), mean squared error (MSE), percentage accuracy (PA) performance metrics. According to the results, all the models performed well and the ARFIMA model performed better than the ARIMA and SARIMA models.

[14] developed an LSTM model for predicting the most significant weather attributes such as precipitation and temperature for a weather station in Sri Lanka. For evaluating model performance, they used Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. As a result, they showed that both LSTM models developed for precipitation and temperature prediction performed well and could be used for making accurate predictions of precipitation and temperature.

[15] developed a SARIMA model using daily temperature time series for Memphis, Tennessee based on data from 2016 to 2019. They analysed temperature data for that period for trends and to see the transient variations over time. They used Mann-Kendall (M-K) test to detect time series analysis patterns as a non-parametric technique. Their study revealed an increase trend of 0.003 °F in temperature for almost every day. They also conducted a forecasting using SARIMA method and estimated temperature values of next 50 days. The forecast also showed and increasing trend for the location.[16] predicted monthly mean minimum and maximum temperatures for the Bhagirathi River basin in India using a seasonal ARIMA model based on data for the period between 2001 and 2020. Their results showed that forecast data fits well with the trend in the data.

[17] developed a SARIMA model for time series prediction of temperature in Pune, India using dataset from year from 2009 to 2020. They used the autocorrelation function, and the partial autocorrelation function as well as using standardized residuals in order to find the best fit for the time series for their study and came up with SARIMA (1, 1, 1) (1, 1, 1) (12) model that could best represent their time series data. They performed model diagnostics, and the model was found to be performing well in predicting temperature values. [18] developed and implemented an LSTM model on AWS machine learning platform. They observed that LSTM model gave substantial results with high accuracy among the other weather forecasting techniques.

2. Methods

A time-series approach was used in this study to predict monthly mean temperature values for a weather station in Ankara, Türkiye. A time series is a sequence of data instances that occur in successive order over period a period of time, such as hourly, weekly, monthly, or yearly. Time series approach makes it possible to develop an appropriate model which describes inherent structures of the series by rigorously analysing and processing data from past observations. This makes it possible to make future predictions, monitoring and control using the developed model [1].

2.1. The study area and Data pre-processing

The data used in this study belong to the Automatic Weather Observation Station (AWOS) which is numbered 17130 and operates as part of Observation Network of the Turkish Meteorological Service and located near the Headquarters of Turkish Meteorological Service in Ankara, Turkey. The location of the meteorological station is shown in Figure 1. It has got geographic coordinates of 39°58'21.0" North and 32°51'50.0 East and the altitude for the station is 883 m. The station is equipped with sensors that measure various meteorological parameters, such as temperature, humidity, precipitation, air pressure, etc. at certain intervals (every second, minute, or hourly, daily, etc.).

The time series data consists of monthly mean temperature readings from the weather station from January 1, 2000, to March 2023 and deemed to represent the local weather in Ankara.

Table 1. Descriptive statistics of temperature data used in the study.

Mean	Std err.	Median	Mod	Std. Dev.	Kurtosis	Skewness	Min.	Max.
12,85	0,50	12.8	11,2	8,4	-1,22	-0,006	-4.0	28,1

The data was obtained as an Excel file in ".xlsx" format and contains a total of 279 monthly mean temperature measurements for each month between Jan. 2020 and Mar. 2023. Table 1 shows the some of the descriptive statistics for the temperature dataset used in the study. The mean minimum and maximum temperature values for the temperature dataset are -4.0 and 28.1 respectively. The kurtosis is -1.22 and the skewness is -0.006 for the distribution of the dataset. The kurtosis value being a negative number indicates that data rather has a flat distribution. Skewness being close to zero means that the tails of the distribution are almost symmetrical and don't differ significantly.

Figure 1. The location of meteorological observation station where the data were obtained from.

A normalization process was carried out before building an LSTM model by scaling all values in the dataset into the range of between 0 and 1. Scaling was not in the case of ARIMA as this is not necessary step for ARIMA models. The dataset then was split into training and testing datasets by reserving 80% of the whole dataset for training and the remaining 20% for the test dataset.

The LSTM and (S)ARIMA models proposed in this study were implemented in Python programming language. A separate Python program was written for developing each of the model. For implementing the SARIMA model, "statsmodels" module, and for implementing the LSTM model the Keras machine learning module was used in the study.

2.2. ARIMA

Autoregressive moving average models are used in cases where the output not only depends on the current information but also all the previous information that have arrived over a previous stretch of time. ARIMA models are only applicable to stationary time series data, therefore it is essential to make sure that the time series data is stationary before it is used in ARIMA models. ARIMA models are generally used for two major purposes: to analyse the data set to better understand it and to predict future trends. Autoregressive statistical models, such as ARIMA predict future values based on past values. A very popular use of ARIMA models is predicting future stock market prices. An ARIMA model could be used to predict prices of a stock future prices based on its past values.

ARIMA models consist of three components: AR, I and MA. The "AR" stands for autoregression, and an autoregressive model is a model with a changing variable that regresses on its own prior or lagged values. Such a model predicts future values based on past values. "I" stands for "integrated" and rep-resents the differencing of time series values to make it stationary. If dataset is not stationary it is made stationary by simply replacing data values with difference between data values and previous values. The last component "MA" stands for Moving Average which is a representation of the error of the model as a linear combination of previous forecast errors [19].

ARIMA models require the time series to be stationary, and non-stationary data cannot be used for building an ARIMA model. Thus, the first step for building a ARIMA model is to make sure that data at hand is stationary. The stationarity of the data used in the study was checked using the Augmented Dickey and Fuller (ADF) unit root test. The ACF test is used in order to check whether or not the data has a unit root, which is stated by the null and alternative hypotheses in the test. The test statistics of the Fuller test verified the stationarity of the time series dataset with p-value smaller than 0.01. Additionally, the ACF plot shown in Figure 2 exhibits a sine wave, which is another evidence that the series is indeed stationary.

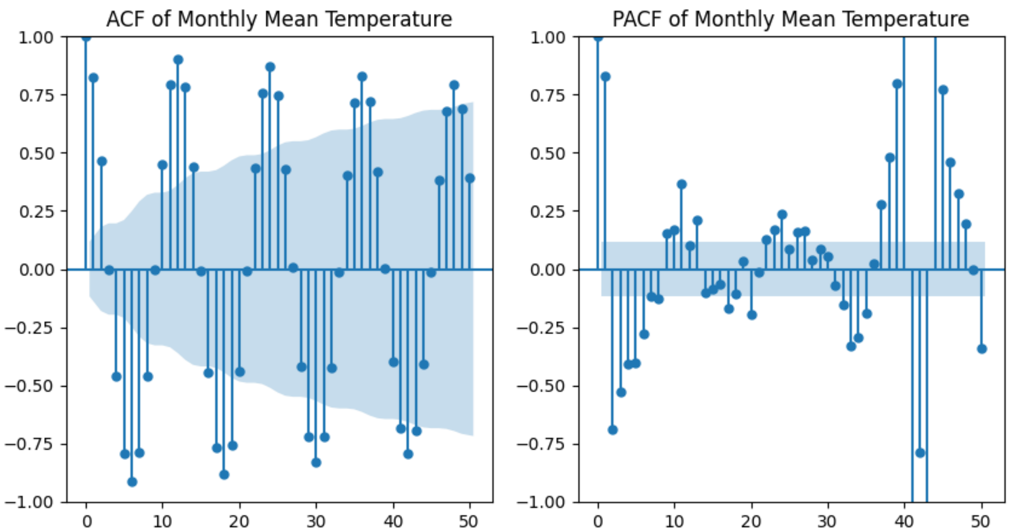

Figure 2. ACF and PACF of monthly mean temperature during Jan. 2000-Mar. 2023. The ACF and PACF graphs were plotted for 50 lags. The ACF graph shows that ACF very slowly tails off towards the end. On the other hand, PACF does not cut off event at lag=50.

As can be seen from the Figure 2, the ACF graph tails off very slowly towards lag=50 while the PACF does not cut off even after lag=50. This suggests that the model is not an AR or a MA model. Thus the "automatic arima" function in a Python module was used in order to find out the best fitting model.

2.3 SARIMA

SARIMA, which is short for Seasonal Autoregressive Integrated Moving Average (also known as Seasonal ARIMA) is basically an ARIMA model, which takes into account seasonality in addition to past values. It has three new parameters to represent seasonal terms in ARIMA for seasonal components as well as an additional parameter for the seasonality period. The new hyperparameters are the seasonal counterparts of those existing in ARIMA and specify autoregression (AR), differencing (I) and moving average (MA). The four seasonal elements that are part of a SARIMA are [12], [20]:

P: Seasonal AR order
D: Seasonal difference order
Q: Seasonal MA order
M: Seasonal period

In order to find out the best performing ARIMA model, automatic arima function, i.e., auto_arima() in "pmdarima" Python module. The results of "auto_arima()" function showed that the dataset has seasonality of 12 months, which was an expected result as monthly temperatures follow a similar pattern throughout the year. The best model for the dataset was found was found as SARIMA (1, 0, 0) × (2, 0, 1) 12 with 12-month seasonality. Since the dataset had seasonality component, this SARIMA model was used in the study for the prediction of temperature values in the test dataset. Residuals were checked for normality and correlation it was found that they exhibit a normal distribution and not correlated as can be seen from the Q-Q plot and the histogram in Figure 3. In order to assess the usability of the model for forecasting a diagnostics test was performed, which showed that the model could be used for prediction.

Table 2. Results of the diagnostics test of the SARIMA (1, 0, 0) × (2, 0, 1) 12 model

	Coef	Std err.	*z*	*P*>|*z*|	[0.025	0.975]
AR.L1	0.3198	0.062	5.147	0.000	0.198	0.442
AR.S.L12	0.7462	0.073	10.220	0.000	0.603	0.889
AR.S.L24	0.2531	0.073	3.476	0.001	0.110	0.396
MA.S.L12	-0.8610	0.060	-14.414	0.000	-0.978	-0.744

Table 2 summarizes the results of SARIMA (1, 0, 0) × (2, 0, 1) 12 model. The 'Coef' column is the weight of coefficients and shows degree of impact each feature on the time series. The fourth column shows the statistical significance of each feature. As can be seen from the values in that column all values are less than 0.05 and therefore statistically significant.

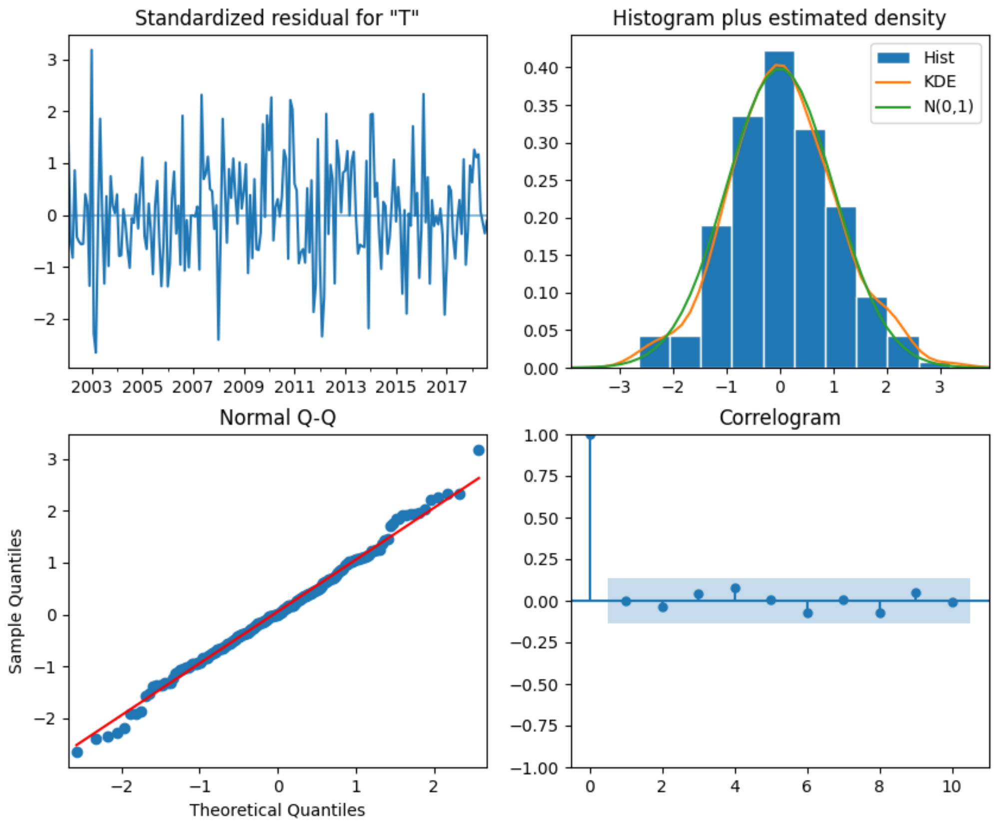

Figure 3. Plots of (a) standardized residual (b) histogram and estimated density (c) Normal Q-Q plot (d) Correlogram

Standard residuals graph shown in Figure 3a appear to be white noise and shows no obvious pattern, suggesting that the model is a good fit for the dataset. The Kernel Density Estimation (KDE) (red curve) shown in Figure 3b seems almost overlapped with the N (0, 1) (green curve), with 0 mean=0 and standard deviation=1. In the Q-Q plot shown in Figure 3c, the red line represents a normal distribution with mean=0 and standard deviation=1 and the blue dots represent residuals. The blue dots being very close to the red dotes showing a normal distribution suggests that the residuals are normally distributed and follow a linear trend. The correlogram shown in Figure 3d the data have a low correlation with the lagged data. All these show that the model fist well to the time series data and can be used for prediction and forecasting.

The selected model was further validated by predicting the monthly mean temperatures in test dataset consisting of temperature values from July 2018 to March 2023 and reproducing the known seasonal patterns in its forecast. This showed that the predicted mean temperature values were the same or very close to actual data and following the same seasonal pattern.

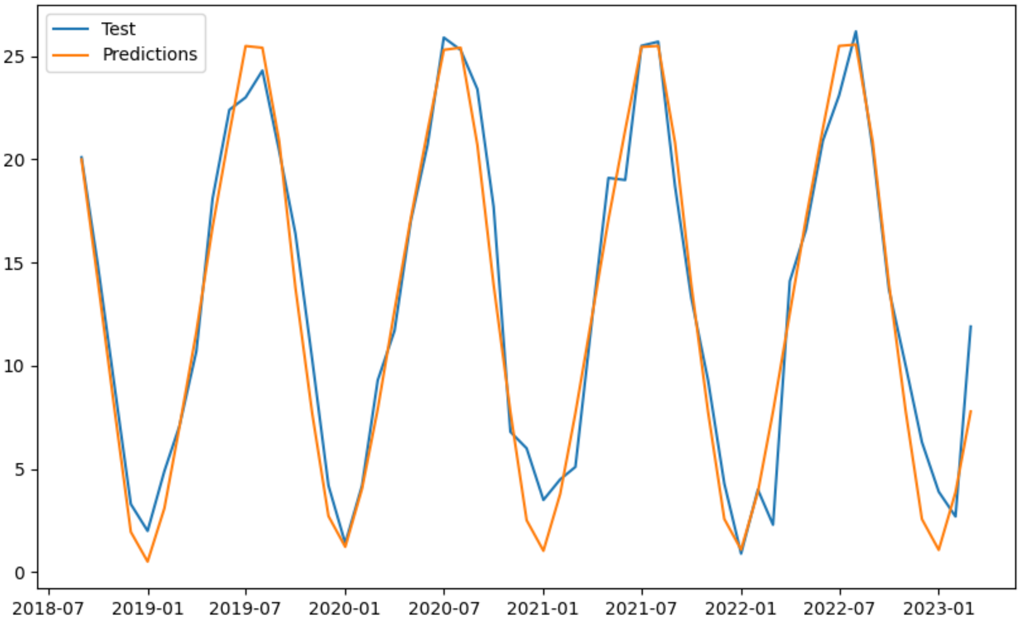

Figure 4. Plot of predicted values together with the actual temperature values suing the SARIMA model.

Figure 4 shows actual and predicted temperature values using the SARIMA model in the same plot. From the plot it can clearly be seen that predicted values by the SARIMA model are very close the actual temperature values and follow a similar pattern.

2.4. Long Short-Term Memory (LSTM)

LSTM networks are a type of recurrent neural network (RNN), which are capable of remembering previous information and using them to learn long-term dependencies. They were first introduced in 1997 by Hochreiter & Schumidhuber and has become a popular methodology for predicting future values based on past data. They produce remarkable good results for a large variety of problems, and mainly used in the field of Deep Learning. The name LSTM refers to the fact that these networks have both short term and long-term memory. The connection weights and biases in the LTSM network change with each episode of training, which constitutes its long-term memory. Likewise, changes in activation patterns with every time-step constitutes its short-term memory. At any particular point in time the output of an LSTM networks depends on three factors, namely current cell state, previous output (hidden state) and the current input [21].

LSTM networks have feedback connections, which makes them different from traditional neural networks. This unique feature of LSTM enables them to retain information and process not just process single data points but also process entire sequence of data, making it suitable for domains like speech recognition, machine translation, image recognition, etc. The ability of LSTM to handle lags between subsequent events make them well-suited for classification, processing and predictions problems based on time series data. A sequential model with only 1 input layer with 7 neurons and a dense output layer of 1 neuron was selected as the best performing model for this study.

A typical LSTM unit is composed of four components: a memory cell, an input gate, an output gate and a forget gate. The memory cell maintains its state over time. The cell state runs down the entire chain with only small linear interactions, enabling the information flow along it unchanged. Information can be added or removed through the gate structures, which perform element-wise multiplication by sigmoid ranges between 0 and 1. The gates open and close and store information in analog format [7], [22].

A thorough hyperparameter sensitivity study was done in order to find the best performing LSTM model for the study. A Python program was written which varied number of neurons in the LSTM layers, number of hidden layers and number of dense layers in for loops and the resulting performances of each were recorded for that purpose. And the best performing model with minimum error rate was selected based on this criterion. The resulting best performing model was a sequential model and consisted only of 1 input layer with 7 neurons and a dense output layer. The LSTM model summary for the study is shown in Figure 5. As can be seen from "Output Shape" column of the figure, the selected LSTM model consist of an input layer with 7 neurons and a dense output layer.

Batch size for the model was set 5 and Adaptive Moment Estimation (ADAM) was used as the optimization algorithm. Mean squared error was used as the loss function of the LSTM algorithm. The number of epochs was set to 10 for training of the model as this yielded least error rate. Increasing number of epochs worsened the performance with increased error rate.

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_2 (LSTM)               (None, 7)                 252

 dense_2 (Dense)             (None, 1)                 8

=================================================================
```
Figure 5. The LSTM model summary

Figure 6 shows the time series plot of actual values and predicted values of the monthly temperature data using the LSTM model developed for the study. As with the ARIMA results above, seasonal variations in the dataset can be clearly seen from the figure. It can also be observed from the figure that although the LSTM model was able to capture the internal structure and follow similar pattern, it underestimates the higher temperatures and overestimates lower temperatures. When compared to the ARIMA results plot in Figure 4, the ARIMA model better approximates the dataset and exhibits closer pattern to the actual data than the LSTM model.

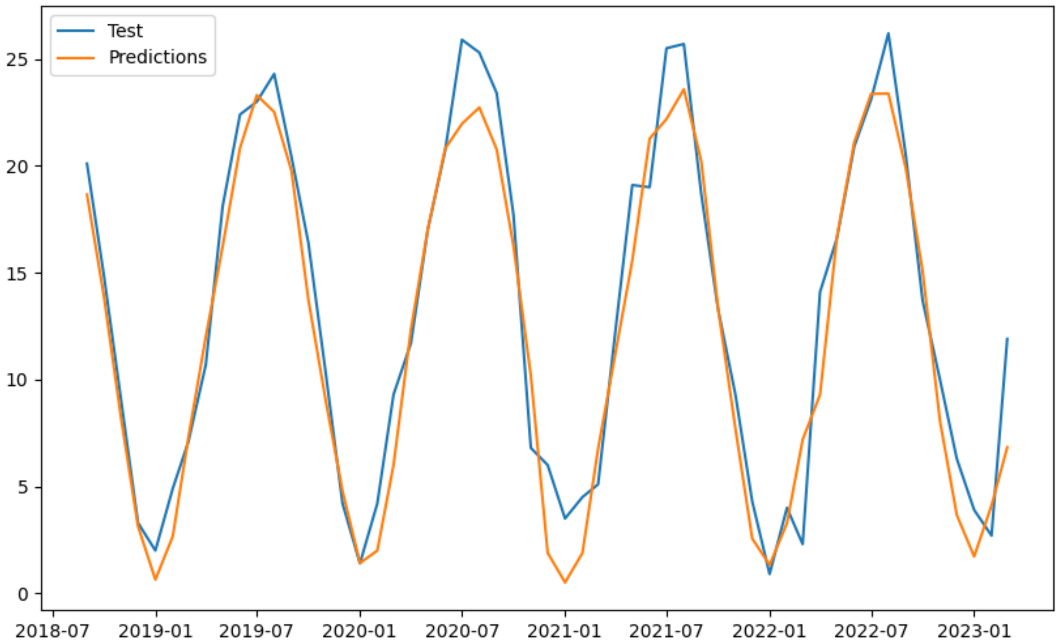

Figure 6. The plot of actual and predicted monthly mean temperature values using the LSTM model.

3. Results

In order to evaluate and compare performances three different performance metrics, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Square Error (MSE) were used in the study.

Root Mean Square is the mean of all prediction errors and calculated using the formula below:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2 \tag{4}$$

RMSE is the root of the MSE, and it is calculated as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2} \tag{5}$$

Mean Absolute Error is the absolute mean of all differences between actual and predicted values and calculated using the formula:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}(|P_i - O_i|) \tag{6}$$

As a result of performance evaluation process, MAE of 1.43 and RMSE of 1.87 for SARIMA model and MAE of 1.69 and RMSE of 2.04 for LSTM model were obtained. The performance results were summarized in Table 3.

Table 3. Results of the performance evaluation of the models

Method	MAE	MSE	RMSE
SARIMA	1.43	3.51	1.87
LSTM	1.69	4.14	2.04

From the Table 3 which summarizes the model results it was concluded that SARIMA model yielded least error prediction of monthly temperatures compared to the LSTM model.

4. Conclusion

This study aimed at predicting monthly temperature values based on historical temperature data for a weather station located in the city of Ankara, Türkiye from January 2010 to March 2023 using two different time series prediction methodologies, namely Seasonal ARIMA (SARIMA) as well as LSTM and their performance results were presented for comparison. The SARIMA and LSTM methodologies used in this study selected based on their popularity and wide use for predicting times series.

As an initial first step, the stationarity of the dataset was tested using the Augmented Dickey and Fuller (ADF) unit root test and the dataset was found to be stationary. This step was necessary as non-stationary data could not be used to build a ARIMA model. The ACF and PACF graphs did not indicate a clear AR, MA or ARIMA model, thus the “automatic arima” function in a Python module was utilized in order to come up with best fitting model. The process revealed the inherent seasonality in the dataset and yielded the model of SARIMA (1, 0, 0) × (2, 0, 1) 12 with 12-month seasonality as the best fitting model to the dataset. Normality check was performed for the normality of residuals, and they were found to be normally distributed and not correlated with each other. Further, a diagnostics test was performed to assess the usability of the model for prediction which verified the usability of the model for that purpose.

The selected model was further validated by predicting the monthly mean temperatures in test dataset consisting of temperature values from July 2018 to March 2023 and reproducing the known seasonal patterns in its forecast. This showed that the predicted mean temperature values were identical or very close to actual data and following the same seasonal pattern.

The selected model then applied to the test dataset consisting of temperature values from July 2018 to March 2023. This showed very close temperature values to the actual values in the test dataset and similar seasonal pattern as could be seen in the plots, validating that the model was able to represent the dataset and could be used as a forecasting tool for future monthly mean temperatures.

For the second method, in which an LSTM network was used, a thorough study was done in order to find the best performing LSTM model by varying number of LSTM, hidden and dense layers as well as number of neurons in each one of them through a Python program and the best performing LSTM model was selected based on the minimum RMSE metric for predicting monthly mean temperatures. The LSTM model was then applied to the test dataset to predict temperature values from July 2010 to March 2023. As with the SARIMA model, LSTM model was also exhibited good fit to the dataset with very close values to the actual temperature values. However, as can be seen from the Figure 6, although the plots of actual and predicted values follow similar pattern, the model mostly underestimates highest values and sometimes overestimates lowest values.

Finally, performance metrics of MSE, RMSE, and MAE were calculated to make prediction performance comparison between for the models. For the SARIMA model the MSE, RMSE and MAE were 1.43, 3.51 and 1.87 and 1.69, 4.14 and 2.04 for the LSTM model respectively. According to these results it can be concluded that both methodologies yielded similar good performances in predicting monthly mean temperatures, but the performance of the SARIMA model was slightly better than that of LSTM model. This could be due to the size of the dataset used for this study as the dataset consisted of only 279 temperature measurements and only 80% (N=223) of the entire dataset was used for building the LSTM model. LSTMs are a special type of Artificial Neural Networks (ANN), thus as with any kind of ANN, training the LSTM model with larger datasets would in general lead to a better model. Therefore, experiments with larger dataset are needed in order to test the superiority of the SARIMA model over the LSTM model in predicting monthly mean temperatures.

Conflict of Interest Statement

The authors declare that there is no conflict of interest.

References

[1] M. Murat, I. Malinowska, M. Gos, J. Krzyszczak, "Forecasting daily meteorological time series using ARIMA and regression models," *International Agrophysics*, vol. 32, no. 2, pp. 253–264, 2018. doi:10.1515/intag-2017-0007

[2] S. Siami-Namini, N. Tavakoli and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," in *Proc. of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 17-20 Dec. 2018, Orlando, FL, USA* [Online]. Available: IEEE Xplore, https://ieeexplore.ieee.org. [Accessed: 7 Agu. 2023]

[3] P. Liu, 'Time Series Forecasting Based on ARIMA and LSTM', in *Proceedings of the 2022 2nd International Conference on Enterprise Management and Economic Development (ICEMED 2022)*, 2022, pp. 1203–1208.

[4] A. Parasyris, G. Alexandrakis, G.V. Kozyrakis, K. Spanoudaki, N.A. Kampanis, "Predicting Meteorological Variables on Local Level with SARIMA, LSTM and Hybrid Techniques", *Atmosphere 2022,* vol. 13, no. 878, 2022. doi:10.3390/atmos13060878

[5] K. Tadesse, O. Megersa, and M. Dinka, "The SARIMA model-based monthly rainfall forecasting for the Turksvygbult Station at the Magoebaskloof Dam in South Africa", *Journal of Water and Land Development*, vol. 53, pp. 100-107, 2022. doi:10.24425/jwld.2022.140785

[6] C. Chatfield, *The analysis of time series: An introduction, 6th ed.* London, UK: Chapman & Hall/CRC, 2004.

[7] I. S. Rahayu, E. C. Djamal, R. Ilyas, "Daily Temperature Prediction Using Recurrent Neural Networks and Long-Short Term Memory", in *Proceedings of the 5th NA International Conference on Industrial Engineering and Operations Management*, *August 10 – 14, 2020, Detroit. Michigan, USA* [Online]. Available: ieomsociety.org, http://www.ieomsociety.org. [Accessed: 10 Apr. 2023].

[8] J. Choi, D. C. Roberts, and E. Lee, "Forecasting oil production in North Dakota using the seasonal autoregressive integrated moving average (S-ARIMA)", *Natural Resources*, vol. 6, no. 1, pp. 16–26, 2015.

[9] A. Can and A. T. Atimtay, "Time series analysis of mean temperature data in Turkey," *Applied Time Series*, vol. 4, no. 20, 2002.

[10] S. Zakaria, N. Al-Ansari, S. Knutsson, and T. Al-Badrany, "ARIMA models for weekly rainfall in the semi-arid Sinjar district at Iraq," *Journal of Earth Science Geotechnical Engineering*, vol. 2, no. 3, pp. 25-55, 2012.

[11] P. Chen, A. Niu, D. Liu, W. Jiang, and B. Ma, "Time series forecasting of temperatures using SARIMA: An example from Nanjing," *IOP Conference Series: Materials Science and Engineering,* vol. 394, 2018. doi:10.1088/1757- 899X/394/5/052024

[12] D. K. Dwivedi, G. R. Sharma and S. S. Wandre, "Forecasting mean temperature using SARIMA Model for Junagadh City of Gujarat," *IJASR*, vol. 7, no. 4, pp. 183–194, Jan. 2017. doi:10.24247/ijasraug201723

[13] J. Asha, S. S. Kumar, and S. Rishidas, 'Forecasting performance comparison of daily maximum temperature using ARMA based methods', *Journal of Physics: Conference Series*, vol. 1921, no. 1, p. 012041, May 2021. doi:10.1088/1742-6596/1921/1/012041

[14] K. M. S. A. Hennayake, R. Dinalankara, and D. Y. Mudunkotuwa, "Machine Learning Based Weather Prediction Model for Short Term Weather Prediction in Sri Lanka," *International Journal of Multidisciplinary Studies (IJMS)*, vol 9, no. I, 2022. doi:10.4038/ijms.v9i1.159

[15] K. N. Mitu, K. Hasan, "Modelling and Forecasting Daily Temperature Time Series in the Memphis, Tennessee," *International Journal of Environmental Monitoring and Analysis*, vol. 9, no. 6, pp. 214-221, Dec. 2021. doi:10.11648/j.ijema.20210906.17

[16] T. Dimri, S. Ahmad, and M. Sharif, "Time series analysis of climate variables using seasonal ARIMA approach," *Journal of Earth System Science volume*, vol. 129, 2020. doi:10.1007/s12040-020-01408-x

[17] A. Gangshetty, G. Kaur, U. S. Malunje, "Time Series Prediction of Temperature in Pune using Seasonal ARIMA Model," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, Issue 11, November-2021.

[18] D. T. Hoang, Pr. L. Yang, L. D. P. Cuong, P. D. Trung, N. H. Tu, L. V. Truong, T. T. Hien, V. T. Nha, "Weather prediction based on LSTM model implemented AWS Machine Learning Platform," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol 8, no. V, May 2020. doi:10.22214/ijraset.2020.5046

[19] "What Is ARIMA Modeling?," *Master's in Data Science*, edX.org, 2023. [Online]. Available: https://www.mastersindatascience.org/learning/statistics-data-science/what-is-arima-modeling. [Accessed: Apr. 10, 2023]

[20] R. J. Hyndman and G. Athanasopoulos, "Forecasting: principles and practice", 2013.

[21] R. Dolphin. "LSTM Networks | A Detailed Explanation | A Comprehensive Introduction to LSTMs". *Towards Data Science,* medium.com, Oct 21, 2020. [Online]. Available: https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9. [Accessed: May 11, 2023].

[22] A. I. Arasu, M. Modani, and N. R. Vadlamani, "Application of Machine Learning Techniques in Temperature Forecast", in *Proc. of the 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2022, Nassau, Bahamas, pp. 513-518, doi:10.1109/ICMLA55696.2022.00083