IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2025) 15 (2): 340-355

UKJAES

University of Kirkuk Journal For Administrative and Economic Science

ISSN:2222-2995 E-ISSN:3079-3521
University of Kirkuk Journal For Administrative and Economic Science

Muhammed Halkawt jamal & Ahmed Nawzad M. . LSTM model performance for Municipal solid waste Time series. *University of Kirkuk Journal For Administrative and Economic Science* (2025) 15 (2):340-355.

LSTM model performance for Municipal solid waste Time series

Halkawt jamal Muhammed [1], Nawzad M. Ahmed [2]

[1,2] *Department of Statistics and Informatics-University of Sulaimani - College of Administration and Economics, Sulaimani, Iraq*

halkawt.muhammad@univsul.edu.iq [1]
nawzad.mahmud@univsul.edu.iq [2]

Abstract: This study evaluates the performance of Long Short-Term Memory (LSTM) models using Tanh, Sigmoid, ReLU, Swish, and Softmax activation functions over 50 epochs for time series prediction. Key performance metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2), were analyzed to assess model accuracy and explanatory power. Results indicate that each activation function exhibits varying levels of predictive performance, with the best R^2 values ranging from 0.5942 to 0.7837 across models. The Tanh-activated LSTM achieved its highest performance at epoch 45 ($R^2 = 0.7837$), the Sigmoid-activated LSTM at epoch 50 ($R^2 = 0.7794$), and the ReLU-activated LSTM showed competitive accuracy in selected epochs. These findings highlight the impact of activation functions on LSTM model efficiency, guiding optimal selection for time series forecasting applications.

Keywords: LSTM, Activation Functions, Time Series Prediction, Model Performance Metrics.

أداء نموذج LSTM للنفايات الصلبة البلدية السلسلة الزمنية

الباحث: هه لكه وت جمال محمد ١، أ.د. نوزاد محمد أحمد ٢

١،٢ قسم الإحصاء والمعلوماتية-جامعة السليمانية-كلية الإدارة والاقتصاد، السليمانية، العراق

المستخلص: تُقيّم هذه الدراسة أداء نماذج الذاكرة طويلة المدى قصيرة المدى (LSTM) باستخدام دوال التنشيط Tanh وSigmoid وReLU وSwish وSoftmax على مدى ٥٠ حقبة زمنية للتنبؤ بالسلاسل الزمنية. وقد حُللت مقاييس الأداء الرئيسية، بما في ذلك جذر متوسط مربع الخطأ (RMSE)، ومتوسط مربع الخطأ (MSE)، ومعيار معلومات أكايكي (AIC)، ومعيار المعلومات البايزي (BIC)، ومعامل R^2، لتقييم دقة النموذج وقدرته التفسيرية. وتشير النتائج إلى أن كل دالة تنشيط تُظهر مستويات متفاوتة من الأداء التنبئي، حيث تتراوح أفضل قيم R^2 بين ٠,٥٩٤٢ و٠,٧٨٣٧ في جميع النماذج. حقق نموذج LSTM المُنشَّط بـ Tanh أعلى أداء له في العصر ٤٥ ($R^2 = 0.7837$)، ونموذج LSTM المُنشَّط بـ Sigmoid في العصر ٥٠ ($R^2 = 0.7794$)، بينما أظهر نموذج LSTM المُنشَّط بـ ReLU دقةً تنافسيةً في عصور مختارة. تُسلِّط هذه النتائج الضوء على تأثير دوال التنشيط على كفاءة نموذج LSTM، مما يُوجِّه الاختيار الأمثل لتطبيقات تنبؤ السلاسل الزمنية.

الكلمات المفتاحية: LSTM، دوال التنشيط، تنبؤ السلاسل الزمنية، مقاييس أداء النموذج.

Corresponding Author: E-mail: halkawt.muhammad@univsul.edu.iq

Introduction

Municipal solid waste (MSW) is one of the most pressing concerns in urban planning today. The rapid acceleration of MSW generation is driven by global urbanization, population growth, and increasing material consumption. As cities expand and undergo transformation, the volume of municipal waste continues to rise, often outpacing the rate of urbanization itself. Managing this growing waste efficiently is a major environmental challenge for metropolitan areas, necessitating proper treatment and sustainable management strategies. Inefficient waste handling in urban settings leads to serious issues, making it difficult to collect, process, and dispose of waste in an environmentally responsible manner,(Huang et al. 2020). Accurately predicting future MSW generation is essential for designing sustainable waste management infrastructure. However, this task is complex due to the influence of various demographic, societal, and economic factors. The relationship between waste generation and these factors is often nonlinear and lacks a clear cause-and-effect pattern. Waste management systems generally focus on two primary functions: waste collection and processing. This study considers the waste collection process.Utku, A., & Kaya, S. K. (2022). To minimize the adverse effects of MSW, an accurate prediction of future waste volumes is a crucial issue. Efficient prediction of MSW generation is vital to the implementation of an optimum MSW management system as a primary tool for MSW management. Fu et al. (2015). Waste quantity predictions serve as the foundation for waste management process adoption, development, and optimization (Cubillos, 2020). Incorrect prediction of waste amounts can adversely affect process costs and cause systems to be inefficient. Another of the main challenges in using prediction models in waste management is the high variability and ambiguity of data wastage. The amount of waste produced can be affected by unpredictable factors such as people's behaviors, holidays, and seasonal conditions. The volatile nature of waste data can cause problems in generating future forecasts. In addition, insufficient and incomplete data will reduce the accuracy of the predictions to be made (Ghanbari et al. 2021). Due to the large quantity of generated waste, an increasing amount of MSW can cause severe damage to the environment and population health (Huang et al. 2020).To address these challenges, machine learning models particularly deep learning techniques are increasingly used for waste prediction. Long Short-Term Memory (LSTM) networks, a type of recurrent neural network (RNN), have shown promising results in capturing long-term dependencies in sequential data, making them well-suited for time series forecasting. In this study, we apply an LSTM model to predict daily waste generation in Sulaimaniyah, Iraq. By leveraging historical waste data, we aim to improve forecasting accuracy, helping urban planners and policymakers develop more efficient waste management strategies. Given the vast quantities of waste produced daily, ineffective MSW management can lead to severe environmental damage and public health Thus, improving waste prediction models is crucial for sustainable urban planning and effective waste management strategies.

1st: Literature Review

The study compares the performance of an LSTM-based deep learning model with other machine learning models to predict garbage volumes in Istanbul. The LSTM model outperforms other models like k-nearest Neighbors, Random Forest, Support Vector Machine, Multi-Layer Perceptron, and Gated Recurrent Unit. Extensive and comparative investigations reveal that the LSTM model is the best approach for estimating waste volumes, making it an ideal tool for waste management in rapidly expanding metropolitan areas. Utku, A., & Kaya, S. K. (2022), The study combines Differential Evolution (DE) optimization with Long Short-Term Memory (LSTM) machine learning to enhance post-disaster garbage collection efficiency. It uses a hybrid technique for dynamic route planning and waste generation forecasts. The approach is tested in a city with frequent earthquakes and sensitivity analysis. Results show it performs better than conventional models, saving time and increasing disaster response effectiveness. This combination provides disaster response teams with a reliable and flexible tool for effective waste management. (Yazdani et al (2024), The study aims to improve trash management by using deep learning models like 1D CNN, LSTM, GRU, and Bi-LSTM for real-time garbage prediction from public bins. Real-time bin data from smart IoT sensors is used instead of historical records. The models' performance is assessed using MAE, MAPE, R2, and RMSE. The LSTM model outperforms the others in trash bin prediction, showing robustness and better accuracy. This makes it a useful tool for waste management planning. (Ahmed, (S., Mubarak, et al (2022), The project aims to improve flood management and disaster response by creating a precise prediction model for flood waste creation using LSTM and SVM. Two datasets are used: one for municipal trash stream materials between 1960 and 2015 and another for items recycled and composted. The study divides waste types into 10 groups, using LSTM and SVM models to predict flood waste production. The study emphasizes the importance of accurate flood waste estimation in disaster management. (Fatovatikhah et al (2024).

2nd: Methodology

1- Long-Short-Term Memory (LSTM) Model [5,7,8,9]

A specific kind of recurrent neural network (RNN) called Long Short-Term Memory (LSTM) is made to efficiently learn from input sequences, especially in situations where long-term dependencies are important. The vanishing gradient problem, which prevents typical RNNs from learning from data with lengthy time intervals between pertinent inputs, is one of the drawbacks of LSTMs, which were created by Sepp Hochreiter and Jürgen Schmidhuber in 1997.

2- LSTM Architecture

Three gates: The input gate, the forget gate, and the output gate control the memory cell in LSTM systems. The information that is added to, removed from, and output from the memory cell is determined by these gates.

A. forget Gate:

The forget gate eliminates data that is no longer relevant in the cell state. The gate receives two inputs, $\mathbf{x_t}$(input at a certain time) and $(\mathbf{h_{t-1}})$(prior cell output), which are multiplied by weight matrices before bias is added. An activation function is applied to the outcome, producing a binary output. A piece of information is lost if the output for a certain cell state is 0, but it is kept for later use if the output is 1. The forget gate's equation is:

The formula: $$f_t = \sigma\left(W_f . \left[\begin{smallmatrix} h_{t-1} \\ x_t \end{smallmatrix}\right] + b_f\right) \quad (1)$$

Here:

- W_f : symbolizes the forget gate's weight matrix.
- $[\mathbf{h_{t-1}}, \mathbf{x_t}]$: indicates that the current input and the prior concealed state have been concatenated.
- $\boldsymbol{b_f}$: Is the Forget Gate biased?
- $\boldsymbol{\sigma}$: is the activation function of the sigmoid.

B. Input Gate

The Input Gate plays a crucial role in updating the cell state by adding new, relevant information. First, a sigmoid function determines how much of the new input should be retained. Then, a tanh function generates a set of candidate values, ranging between -1 and 1, based on the current input (x_t) and the previous hidden state $(\mathbf{h_{t-1}})$. Finally, the selected new information is incorporated into the memory by multiplying the filtered values with the generated candidate values. This ensures that only meaningful data is added while irrelevant details are discarded.

The formula : $i_t = \sigma\left(W_t \cdot \left[\begin{smallmatrix} h_{t-1} \\ x_t \end{smallmatrix}\right] + b_i\right)$ (2)

$$\hat{c}_t = tanh(W_c \cdot \begin{bmatrix} h_{t-1} \\ x_t \end{bmatrix} + b_c \qquad (3)$$

We begin with the prior state and eliminate the data that was previously thought to be superfluous. After correcting for the required modifications in each state value, we next add (the formula c_t), which denotes the revised candidate values.

The formula $c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t$ (4)

where

- $\odot$ represents multiplication of elements.
- Tanh is the activation function for tanh.

C. Output gate:

In order to produce the output, the Output Gate must first extract pertinent information from the current cell state. A tanh function** that scales the cell state's values between -1 and 1 is applied first. A sigmoid function then uses the current input (x_t)) and the prior hidden state (h_{t-1})) to calculate how much of this modified information should be kept. The next hidden state is then created by multiplying the filtered data by the converted cell state. In order to guarantee that only significant information is transmitted throughout the sequence, this updated concealed state is subsequently passed forward as the input and output for the subsequent time step.

The formula $o_t = \sigma\left(W_o . \left[\begin{smallmatrix} h_{t-1} \\ x_t \end{smallmatrix}\right] + b_o\right)$ (5)

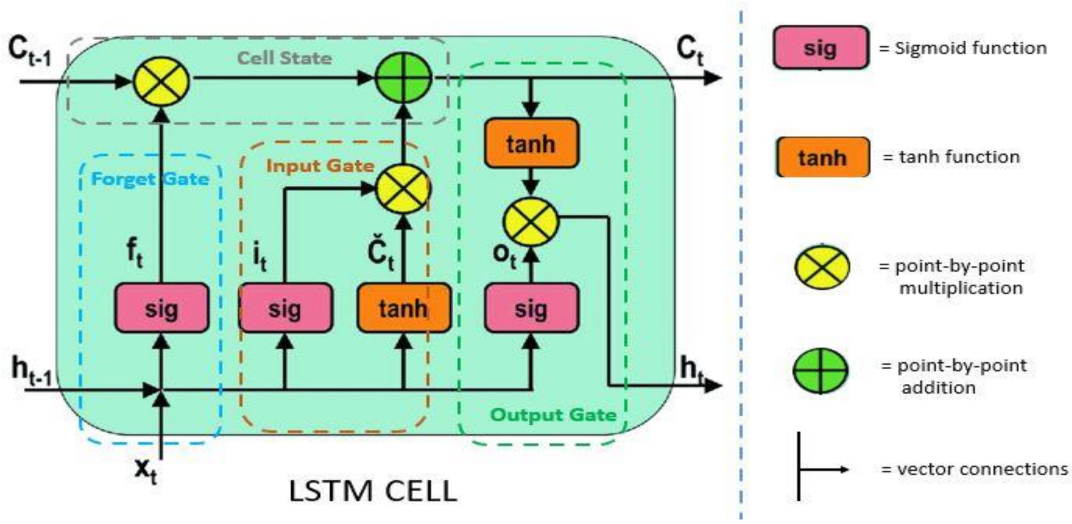

Fig. 1. Structure of LSTM (19)

3- activation function in LSTM

In an LSTM model, different activation functions can be applied at different parts of the network, such as the gates (input, forget, and output) and the hidden state. Below is an example of an LSTM model in TensorFlow/Keras that utilizes five different activation functions:

a. Sigmoid: The Sigmoid Activation Function, a mathematical formula with an 'S' shape, is crucial for gradient-based optimization methods and neural networks. It ensures smooth, continuous output, allowing for complex patterns. Its output ranges between 0 and 1.

The formula: $\sigma(x) = \frac{1}{1+e^{-x}}$ range (0,1) (6)

b. Tanh: The Tanh function is a hyperbolic tangent function, a shifted version of the sigmoid, that can stretch across the y-axis and output values from -1 to +1. It is non-linear, enabling complex data pattern modeling, and is commonly used in hidden layers for easier learning.

The formula: $\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-}}$ Range:(-1,1) (7)

c. ReLU: The ReLU function is a non-linear activation function that improves neural network learning efficiency and reduces computational costs compared to tanh and sigmoid, with a range of [0,∞) for input.

The formula: RELU(x)= max (0 , x) Range: (0,∞) (8)

d. Swish: the Swish function discovered by Google researchers, is a nit Monotonic activation function that can decrease in value despite increasing input values. It can outperform ReLU functions in some cases and is mathematically expressed as:

The formula: $swish(x) = x.\frac{1}{1+e^{-x}}$ Range: $(-\infty,\infty)$ (9)

e. SoftMax: The Softmax activation function is a multiclass classification method that uses multiple sigmoid functions to calculate the probability of each data point in all individual classes, unlike binary classification.

The formula: $\text{SoftMax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_i}}$ **Range:** (0,1) (10)

3rd: Procedures of the LSTM Model

Several stages need to be considered to yield an accurate tourism demand forecasting model using LSTM. All analyses will be performed using Python software.

1- Stage 1: Data preparation

The raw historical data obtained from relevant resources will undergo this process to ensure that the data are clean and compatible with the LSTM model.

2- Stage 2: Data normalization

This process will re-scale the data within the range [0,1] by using MinMaxScaler. This process is required to enhance the learning acceleration rate and at the same time to improve the forecast accuracy level.

3- Stage 3: Data splitting

The data will be partitioned into two parts; an in-sample data set (training data set) and an out-sample data set (testing data set) based on ratio 80:20 respectively. The training data set will be inserted into the training algorithm to identify the output of our prediction model, meanwhile, the testing data set will perform a reasonable check on the algorithm.

4- Stage 4: Network training

Throughout this process, the hyper-parameters known as relevant parameters that control the learning process will be adjusted to enhance the accuracy level and at the same time avoid overfitting or under-fitting. The hyper-parameters are listed as follows: number of nodes or hidden layers; the number of units in the dense layer; activation function; a number of training epochs; batch of size; learning and decay rate; dropout layer; weight initialization; learning rate and momentum. Generally, there are no rules or guidelines on how to identify the value of hyper-parameters. However, we can use a trial and error approach to figure out the value of hyper-parameters, depending on the complexity of the data.

5- Stage 5: Forecasting Finally,

the testing data will be inserted into a trained model for forecasting. The denormalization procedures will be performed on the obtained prediction data afterward. The summary of LSTM procedures in demand forecasting.

4th: Evaluate Precision of Forecasting Models

Evaluating the precision of forecasting models is a critical step in determining how well a model predicts future values. The precision of a forecasting model refers to its ability to generate accurate, reliable, and consistent predictions. There are various statistical measures and methods used to assess the precision of forecasting models, depending on the type of model being used (e.g., time series, regression, machine learning). Below is an overview of some common methods and metrics used to evaluate the precision of forecasting models:

1- Mean Squared Error (MSE)

MSE is a metric used to evaluate how well a model's predictions match the actual data. It calculates the average of the squared differences between the predicted values and the actual values.

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2 \qquad (11)$$

where:

- n is the number of observations.
- y_t is the actual value.
- $\hat{y}_t$ is the predicted value.

2- Root Mean Squared Error (RMSE)

The **Root Mean Squared Error (RMSE)** is another commonly used metric to evaluate forecasting accuracy. RMSE penalizes large errors more heavily than MAE because it squares the differences between actual and forecasted values before averaging them. This makes RMSE particularly sensitive to outliers and large deviations in forecasts.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \qquad (12)$$

3- Akaike Information Criterion (AIC)

AIC is a model selection criterion that helps evaluate how well a statistical model fits the data while penalizing the model for having too many parameters (complexity). It is widely used to compare different models.

$$AIC = 2\,k - 2\,\ln(\mathscr{L}) \qquad (13)$$

Where:

- n: is the number of observations.
- $\mathscr{L}$: is the log-likelihood.

4- Bayesian Information Criterion (BIC)

BIC is very similar to AIC in that it also balances model fit with model complexity, but it applies a stronger penalty for complexity. It is based on Bayesian principles and is often used in statistical model selection.

$$BIC = k \cdot \ln(n) - 2\ln(\mathscr{L}) \qquad (14)$$

Where:

- n: is the number of observations.
- $\mathscr{L}$: is the log-likelyhood.
- k: is the number of explanatory variables in the model.

5th: Results and Discussion:

1- Data Collection

The historical data that will be considered in this study is the amount of waste received daily(kg) from December 2022 until October 2024(654 data). The data were taken from the company Ecocem of Sulaymaniyh.

2- Analysis of Results:

The performance of Long Short-Term Memory (LSTM) models using Tanh, Sigmoid, ReLU, Swish, and Softmax activation functions over 50 epochs for time series prediction. Key performance metrics, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2):

Table (1): LSTM (Tanh)

LSTM (Tanh)					
Epoch	**RMSE**	**MSE**	**AIC**	**BIC**	**R-square**
1	60609.47	3673507975	14427.98	14432.4649	0.7313
2	57774.34	3337873796	14365.23	14369.70994	0.7559
3	72280.31	5224443922	14658.68	14663.16123	0.6179
4	69182.43	4786209146	14601.29	14605.77438	0.6500
5	58426.49	3413654173	14379.93	14384.41096	0.7503
6	61848.42	3825226710	14454.49	14458.97454	0.7202
7	56837.38	3230487527	14343.81	14348.2897	0.7637
8	69064.48	4769901707	14599.05	14603.53808	0.6512
9	57123.53	3263097223	14350.38	14354.86501	0.7614
10	66015.56	4358054611	14539.9	14544.38946	0.6813
11	63433.47	4023805687	14487.64	14492.12339	0.7057
12	56893.52	3236872800	14345.1	14349.57978	0.7633
13	64565.52	4168706502	14510.81	14515.29498	0.6951
14	65757.5	4324048333	14534.78	14539.25968	0.6838
15	56625.34	3206428847	14338.91	14343.39432	0.7655
16	58549.34	3428025648	14382.68	14387.1659	0.7493
17	73277.5	5369591567	14676.62	14681.10733	0.6073
18	72678.32	5282138852	14665.87	14670.35478	0.6137
19	61835.35	3823609891	14454.21	14458.69916	0.7204
20	64897.48	4211682443	14517.53	14522.01388	0.6920
21	63445.29	4025304823	14487.89	14492.37119	0.7056
22	64358.36	4141998746	14506.6	14511.08828	0.6971

23	58901.36	3469370787	14390.53	14395.01812	0.7463
24	58627.36	3437167833	14384.43	14388.90995	0.7486
25	67820.54	4599625714	14575.24	14579.7269	0.6636
26	71845.4	5161762090	14650.77	14655.25349	0.6225
27	70313.35	4943967385	14622.53	14627.01733	0.6384
28	70552.51	4977656498	14626.98	14631.46259	0.6360
29	68377.5	4675482411	14585.96	14590.44188	0.6581
30	60903.44	3709228870	14434.32	14438.80404	0.7287
31	60391.47	3647129552	14423.26	14427.74456	0.7333
32	58180.33	3384951136	14374.4	14378.88362	0.7524
33	73563.53	5411593123	14681.73	14686.21031	0.6042
34	59723.48	3566893896	14408.69	14413.17359	0.7391
35	73394.56	5386761834	14678.71	14683.19732	0.6060
36	67561.35	4564536541	14570.23	14574.71452	0.6662
37	63344.3	4012500051	14485.8	14490.2841	0.7065
38	59137.36	3497227229	14395.77	14400.25644	0.7442
39	67792.35	4595802529	14574.7	14579.18594	0.6639
40	70999.41	5040916121	14635.25	14639.73624	0.6313
41	67466.54	4551733588	14568.39	14572.87119	0.6671
42	66323.5	4398806918	14546	14550.48719	0.6783
43	65006.55	4225851556	14519.73	14524.21229	0.6909
44	58649.37	3439748226	14384.92	14389.40144	0.7484
45	54381.77	2957377380	14285.96	14290.442	0.7837
46	58908.5	3470211160	14390.69	14395.1738	0.7462
47	57548.52	3311832511	14360.09	14364.57544	0.7578
48	68599.43	4705882427	14590.2	14594.68817	0.6558
49	66479.47	4419520118	14549.08	14553.56485	0.6768
50	62323.34	3884198110	14464.51	14468.99706	0.7159

The table1: provides performance metrics for an LSTM (Long Short-Term Memory) model with a Tanh activation function over 50 epochs, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). RMSE and MSE measure the model's prediction errors, with lower values indicating better performance, while AIC and BIC are used for model selection, balancing goodness-of-fit and complexity, with lower values suggesting a better model. R^2, ranging from 0 to 1, represents the proportion of variance in the target variable explained by the model, with higher values indicating better explanatory power. Over the epochs, the model shows variability in performance, with R^2 values fluctuating between 0.6042 and 0.7837, indicating periods of stronger and weaker predictive accuracy. The best performance is observed at epoch 45, with the highest R^2 (0.7837) and the lowest RMSE (54381.77434), suggesting optimal model fit at this stage.

Table (2): LSTM (Sigmoid)

LSTM (Sigmoid)					
Epoch	RMSE	MSE	AIC	BIC	R-square
1	61181.47	3.74E+09	14440.3	14444.78	0.7262
2	58236.84	3.39E+09	14375.68	14380.16	0.7520
3	72730.31	5.29E+09	14666.81	14671.3	0.6131
4	69719.93	4.86E+09	14611.44	14615.92	0.6445
5	58988.99	3.48E+09	14392.49	14396.97	0.7455

6	62360.92	3.89E+09	14465.31	14469.79	0.7156
7	57524.88	3.31E+09	14359.56	14364.05	0.7580
8	69701.98	4.86E+09	14611.1	14615.58	0.6447
9	57811.03	3.34E+09	14366.06	14370.55	0.7556
10	66415.56	4.41E+09	14547.83	14552.31	0.6774
11	64120.97	4.11E+09	14501.77	14506.25	0.6993
12	57418.52	3.3E+09	14357.14	14361.62	0.7589
13	65153.02	4.24E+09	14522.69	14527.17	0.6896
14	66245	4.39E+09	14544.46	14548.95	0.6791
15	57125.34	3.26E+09	14350.43	14354.92	0.7613
16	59011.84	3.48E+09	14393	14397.48	0.7453
17	73890	5.46E+09	14687.54	14692.02	0.6007
18	73165.82	5.35E+09	14674.63	14679.12	0.6085
19	62485.35	3.9E+09	14467.92	14472.41	0.7145
20	65559.98	4.3E+09	14530.84	14535.33	0.6857
21	63932.79	4.09E+09	14497.92	14502.4	0.7011
22	65020.86	4.23E+09	14520.03	14524.51	0.6908
23	59551.36	3.55E+09	14404.92	14409.4	0.7406
24	59214.86	3.51E+09	14397.5	14401.98	0.7436
25	68283.04	4.66E+09	14584.16	14588.64	0.6590
26	72495.4	5.26E+09	14662.57	14667.06	0.6156
27	70913.35	5.03E+09	14633.67	14638.15	0.6322
28	71077.51	5.05E+09	14636.7	14641.18	0.6305
29	68977.5	4.76E+09	14597.41	14601.9	0.6520
30	61503.44	3.78E+09	14447.17	14451.66	0.7234
31	60916.47	3.71E+09	14434.61	14439.09	0.7286
32	58617.83	3.44E+09	14384.22	14388.71	0.7487
33	74226.03	5.51E+09	14693.48	14697.96	0.5971
34	60185.98	3.62E+09	14418.81	14423.29	0.7351
35	73932.06	5.47E+09	14688.28	14692.77	0.6003
36	68223.85	4.65E+09	14583.02	14587.5	0.6596
37	63794.3	4.07E+09	14495.08	14499.56	0.7024
38	59562.36	3.55E+09	14405.16	14409.65	0.7405
39	68229.85	4.66E+09	14583.14	14587.62	0.6595
40	71674.41	5.14E+09	14647.65	14652.14	0.6243
41	67904.04	4.61E+09	14576.86	14581.35	0.6628
42	66911	4.48E+09	14557.57	14562.05	0.6726
43	65694.05	4.32E+09	14533.52	14538.01	0.6844
44	59124.37	3.5E+09	14395.49	14399.98	0.7443
45	62723.34	3.93E+09	14472.9	14477.39	0.7123
46	59358.5	3.52E+09	14400.67	14405.15	0.7423
47	58198.52	3.39E+09	14374.82	14379.3	0.7523
48	69174.43	4.79E+09	14601.15	14605.63	0.6500
49	67041.97	4.49E+09	14560.13	14564.61	0.6713
50	54919.27	3.02E+09	14298.84	14303.33	0.7794

Table 2: presents performance metrics for an LSTM model using the Sigmoid activation function over 50 epochs, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). RMSE and MSE measure prediction errors, with lower values indicating better performance, while AIC and BIC evaluate model fit, balancing complexity and accuracy, with lower values being preferable. R^2, ranging from 0 to 1, indicates the proportion of variance in the target variable explained by the model, with higher values reflecting better predictive power. The model's performance varies across epochs, with R^2 values fluctuating between 0.5971 and 0.7794, showing periods of stronger and weaker accuracy. The best performance occurs at epoch 50, with the highest R^2 (0.7794) and the lowest RMSE (54919.27434), suggesting optimal model fit at this stage. Overall, the Sigmoid-activated LSTM demonstrates competitive performance, with some epochs outperforming others in terms of error reduction and explanatory power.

Table (3): LSTM (RelU)

LSTM (RelU)					
Epoch	**RMSE**	**MSE**	**AIC**	**BIC**	**R-square**
1	62317.9	3.88E+09	14464.41	14468.89	0.7160
2	59475.17	3.54E+09	14403.24	14407.73	0.7413
3	73707.8	5.43E+09	14684.3	14688.79	0.6027
4	70678.26	5E+09	14629.32	14633.8	0.6347
5	59755.26	3.57E+09	14409.4	14413.88	0.7389
6	63450.55	4.03E+09	14488	14492.49	0.7056
7	58735.63	3.45E+09	14386.85	14391.34	0.7477
8	70104.72	4.91E+09	14618.65	14623.13	0.6406
9	58421.93	3.41E+09	14379.84	14384.32	0.7504
10	67610.1	4.57E+09	14571.18	14575.67	0.6657
11	65278.58	4.26E+09	14525.21	14529.69	0.6884
12	57880.61	3.35E+09	14367.64	14372.13	0.7550
13	66273.95	4.39E+09	14545.03	14549.52	0.6788
14	66934.42	4.48E+09	14558.02	14562.51	0.6723
15	58523.59	3.43E+09	14382.11	14386.6	0.7495
16	60219.81	3.63E+09	14419.54	14424.03	0.7348
17	74492.38	5.55E+09	14698.17	14702.66	0.5942
18	73824.87	5.45E+09	14686.38	14690.87	0.6014
19	63323.58	4.01E+09	14485.38	14489.86	0.7067
20	65922.53	4.35E+09	14538.07	14542.55	0.6822
21	64827.22	4.2E+09	14516.12	14520.6	0.6926
22	66021.23	4.36E+09	14540.03	14544.51	0.6812
23	60488.31	3.66E+09	14425.37	14429.85	0.7324
24	60032.07	3.6E+09	14415.45	14419.94	0.7364
25	68777.26	4.73E+09	14593.6	14598.09	0.6541
26	73272.89	5.37E+09	14676.55	14681.03	0.6073
27	72204.01	5.21E+09	14657.3	14661.78	0.6187
28	71987.59	5.18E+09	14653.37	14657.85	0.6210
29	70002.41	4.9E+09	14616.73	14621.22	0.6416
30	62406.86	3.89E+09	14466.27	14470.76	0.7152
31	61378.56	3.77E+09	14444.51	14448.99	0.7245
32	59744.5	3.57E+09	14409.16	14413.65	0.7390

33	74641.74	5.57E+09	14700.8	14705.28	0.5925
34	61295.23	3.76E+09	14442.73	14447.21	0.7252
35	74935.95	5.62E+09	14705.95	14710.43	0.5893
36	69292.56	4.8E+09	14603.38	14607.87	0.6488
37	64566.77	4.17E+09	14510.85	14515.33	0.6951
38	60458.55	3.66E+09	14424.72	14429.21	0.7327
39	68802.22	4.73E+09	14594.08	14598.56	0.6538
40	72426.9	5.25E+09	14661.34	14665.82	0.6164
41	68985.14	4.76E+09	14597.56	14602.04	0.6520
42	67257.44	4.52E+09	14564.33	14568.82	0.6692
43	66464.41	4.42E+09	14548.79	14553.28	0.6769
44	59894.62	3.59E+09	14412.45	14416.93	0.7376
45	58846.93	3.46E+09	14389.33	14393.82	0.7467
46	59971.52	3.6E+09	14414.13	14418.61	0.7370
47	56082.61	3.15E+09	14326.3	14330.79	0.7700
48	70080.07	4.91E+09	14618.19	14622.67	0.6408
49	67952.52	4.62E+09	14577.8	14582.28	0.6623
50	63781.2	4.07E+09	14494.81	14499.29	0.7025

Table 3: provides performance metrics for an LSTM model using the ReLU activation function over 50 epochs, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). RMSE and MSE measure prediction errors, with lower values indicating better performance, while AIC and BIC evaluate model fit, balancing complexity and accuracy, with lower values being preferable. R^2, ranging from 0 to 1, indicates the proportion of variance in the target variable explained by the model, with higher values reflecting better predictive power. The model's performance varies across epochs, with R^2 values fluctuating between 0.5893 and 0.7700, showing periods of stronger and weaker accuracy. The best performance occurs at epoch 47, with the highest R^2 (0.7700) and the lowest RMSE (56082.61155), suggesting optimal model fit at this stage. Overall, the ReLU-activated LSTM demonstrates competitive performance, with some epochs outperforming others in terms of error reduction and explanatory power.

Table (4): LSTM (Swish)

LSTM (Swish)					
Epoch	**RMSE**	**MSE**	**AIC**	**BIC**	**R-square**
1	61844.94	3.82E+09	14454.43	14458.91	0.7203
2	58633.79	3.44E+09	14384.58	14389.06	0.7486
3	73567.64	5.41E+09	14681.81	14686.29	0.6042
4	70156.73	4.92E+09	14619.62	14624.1	0.6400
5	59226.67	3.51E+09	14397.76	14402.24	0.7435
6	63100.55	3.98E+09	14480.76	14485.24	0.7088
7	57711.65	3.33E+09	14363.81	14368.3	0.7564
8	69775.75	4.87E+09	14612.48	14616.97	0.6439
9	58094.12	3.37E+09	14372.46	14376.95	0.7532
10	66649.04	4.44E+09	14552.43	14556.91	0.6751
11	64470.75	4.16E+09	14508.9	14513.38	0.6960
12	57567.75	3.31E+09	14360.54	14365.02	0.7576
13	65951.02	4.35E+09	14538.63	14543.12	0.6819
14	66450.25	4.42E+09	14548.51	14553	0.6771

15	57510.72	3.31E+09	14359.24	14363.73	0.7581
16	59912.62	3.59E+09	14412.84	14417.33	0.7375
17	74629.66	5.57E+09	14700.58	14705.07	0.5927
18	73804.51	5.45E+09	14686.02	14690.5	0.6016
19	63111.56	3.98E+09	14480.98	14485.47	0.7087
20	65812.5	4.33E+09	14535.88	14540.36	0.6832
21	64352.9	4.14E+09	14506.5	14510.98	0.6971
22	65054.82	4.23E+09	14520.71	14525.2	0.6905
23	59564.48	3.55E+09	14405.21	14409.69	0.7405
24	59347.9	3.52E+09	14400.44	14404.92	0.7424
25	68420.68	4.68E+09	14586.79	14591.28	0.6576
26	73027.15	5.33E+09	14672.15	14676.63	0.6100
27	71508.07	5.11E+09	14644.61	14649.1	0.6260
28	71899.11	5.17E+09	14651.76	14656.24	0.6219
29	69166.57	4.78E+09	14601	14605.48	0.6501
30	61566.55	3.79E+09	14448.51	14453	0.7228
31	61460.23	3.78E+09	14446.25	14450.74	0.7237
32	59126.84	3.5E+09	14395.55	14400.03	0.7443
33	74498.93	5.55E+09	14698.29	14702.77	0.5941
34	60888.56	3.71E+09	14434.01	14438.49	0.7289
35	74705.97	5.58E+09	14701.92	14706.41	0.5918
36	68402.29	4.68E+09	14586.44	14590.93	0.6578
37	63998.15	4.1E+09	14499.26	14503.74	0.7005
38	60419.13	3.65E+09	14423.87	14428.35	0.7330
39	68959.28	4.76E+09	14597.07	14601.55	0.6522
40	55287.54	3.06E+09	14307.6	14312.08	0.7765
41	68446.39	4.68E+09	14587.29	14591.77	0.6574
42	67140.35	4.51E+09	14562.05	14566.53	0.6703
43	66236.46	4.39E+09	14544.29	14548.78	0.6791
44	59323.59	3.52E+09	14399.9	14404.38	0.7426
45	71916.28	5.17E+09	14652.07	14656.55	0.6218
46	59819.82	3.58E+09	14410.81	14415.3	0.7383
47	58272.76	3.4E+09	14376.49	14380.97	0.7517
48	69164.38	4.78E+09	14600.96	14605.44	0.6501
49	67098.13	4.5E+09	14561.22	14565.71	0.6707
50	63697.72	4.06E+09	14493.09	14497.58	0.7033

Table 4: presents performance metrics for an LSTM model using the Swish activation function over 50 epochs, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). RMSE and MSE measure prediction errors, with lower values indicating better performance, while AIC and BIC evaluate model fit, balancing complexity and accuracy, with lower values being preferable. R^2, ranging from 0 to 1, indicates the proportion of variance in the target variable explained by the model, with higher values reflecting better predictive power. The model's performance varies across epochs, with R^2 values fluctuating between 0.5918 and 0.7765, showing periods of stronger and weaker accuracy. The best performance occurs at epoch 40, with the highest R^2 (0.7765) and the lowest RMSE (55287.5357), suggesting optimal model fit at this stage. Overall, the Swish-activated LSTM demonstrates competitive performance, with some epochs outperforming others in terms of error reduction and explanatory power.

Table (5): LSTM (Soft max)

LSTM (Softmax)					
Epoch	**RMSE**	**MSE**	**AIC**	**BIC**	**R-square**
1	62366.14	3.89E+09	14465.42	14469.9	0.7155
2	59804.09	3.58E+09	14410.47	14414.95	0.7384
3	73850.56	5.45E+09	14686.84	14691.32	0.6011
4	70476.96	4.97E+09	14625.58	14630.07	0.6367
5	60246.18	3.63E+09	14420.11	14424.6	0.7346
6	63108.81	3.98E+09	14480.93	14485.41	0.7087
7	59163.85	3.5E+09	14396.37	14400.85	0.7440
8	70618.96	4.99E+09	14628.22	14632.71	0.6353
9	58825.06	3.46E+09	14388.84	14393.33	0.7469
10	68310.53	4.67E+09	14584.68	14589.17	0.6587
11	64533.69	4.16E+09	14510.17	14514.66	0.6954
12	58463.76	3.42E+09	14380.77	14385.26	0.7500
13	66311.69	4.4E+09	14545.78	14550.26	0.6784
14	67283.1	4.53E+09	14564.83	14569.31	0.6689
15	58347.88	3.4E+09	14378.17	14382.66	0.7510
16	60279.76	3.63E+09	14420.84	14425.33	0.7343
17	74797.85	5.59E+09	14703.53	14708.02	0.5908
18	74051.63	5.48E+09	14690.4	14694.88	0.5990
19	63444.97	4.03E+09	14487.89	14492.37	0.7056
20	66412.57	4.41E+09	14547.77	14552.26	0.6774
21	64960.39	4.22E+09	14518.81	14523.29	0.6914
22	65999.5	4.36E+09	14539.6	14544.08	0.6814
23	60550.38	3.67E+09	14426.71	14431.2	0.7319
24	60061.06	3.61E+09	14416.08	14420.57	0.7362
25	68957.52	4.76E+09	14597.03	14601.52	0.6522
26	73376.26	5.38E+09	14678.4	14682.88	0.6062
27	71684.03	5.14E+09	14647.83	14652.32	0.6242
28	71839.16	5.16E+09	14650.66	14655.15	0.6226
29	70465.03	4.97E+09	14625.36	14629.85	0.6369
30	62807.16	3.94E+09	14474.65	14479.13	0.7115
31	61914.45	3.83E+09	14455.9	14460.38	0.7196
32	60391.27	3.65E+09	14423.27	14427.75	0.7333
33	75498.76	5.7E+09	14715.75	14720.24	0.5831
34	61325.23	3.76E+09	14443.37	14447.85	0.7250
35	75195.88	5.65E+09	14710.49	14714.97	0.5865
36	68664.2	4.71E+09	14591.45	14595.93	0.6552
37	65106.22	4.24E+09	14521.75	14526.23	0.6900
38	60626.2	3.68E+09	14428.35	14432.84	0.7312
39	69906.13	4.89E+09	14614.93	14619.42	0.6426
40	72496.13	5.26E+09	14662.59	14667.07	0.6156
41	69094.55	4.77E+09	14599.63	14604.12	0.6509
42	68261.36	4.66E+09	14583.74	14588.22	0.6592
43	66776.35	4.46E+09	14554.93	14559.41	0.6739
44	60920.7	3.71E+09	14434.7	14439.18	0.7286
45	67963.06	4.62E+09	14578	14582.49	0.6622
46	59958.83	3.6E+09	14413.85	14418.34	0.7371
47	58688.13	3.44E+09	14385.79	14390.28	0.7481
48	70048.89	4.91E+09	14617.6	14622.09	0.6411
49	55802.34	3.11E+09	14319.74	14324.22	0.7723
50	64258.56	4.13E+09	14504.58	14509.06	0.6980

Table 5: provides performance metrics for an LSTM model using the Softmax activation function over 50 epochs, including Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). RMSE and MSE measure prediction errors, with lower values indicating better performance, while AIC and BIC evaluate model fit, balancing complexity and accuracy, with lower values being preferable. R^2, ranging from 0 to 1, indicates the proportion of variance in the target variable explained by the model, with higher values reflecting better predictive power. The model's performance varies across epochs, with R^2 values fluctuating between 0.5831 and 0.7723, showing periods of stronger and weaker accuracy. The best performance occurs at epoch 49, with the highest R^2 (0.7723) and the lowest RMSE (55802.34327), suggesting optimal model fit at this stage. Overall, the Softmax-activated LSTM demonstrates competitive performance, with some epochs outperforming others in terms of error reduction and explanatory power.

Table (6): Testing (LSTM)

Testing (LSTM)						
Activation	**Epoch**	**RMSE**	**MSE**	**AIC**	**BIC**	**R-Square**
Tanh	45	54147.41	2931942174.32	14280.30	14284.78	0.7856
Sigmoid	50	54988.43	3023727886.44	14300.49	14304.98	0.7789
RelU	47	56032.16	3139602399.04	14325.12	14329.61	0.7704
Swish	40	55367.70	3065582733.54	14309.49	14313.98	0.7758
Softmax	49	55800.69	3113716939.64	14319.70	14324.18	0.7723

Table 6: compares the testing performance of LSTM models with different activation functions (Tanh, Sigmoid, ReLU, Swish, and Softmax) based on their best-performing epochs. The metrics include Root Mean Squared Error (RMSE), Mean Squared Error (MSE), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and R-squared (R^2). The Tanh activation function achieves the best performance with the lowest RMSE (54147.41), lowest MSE (2931942174.32), lowest AIC (14280.30), lowest BIC (14284.78), and the highest R^2 (0.7856), indicating it explains the most variance in the target variable with the least error. Sigmoid follows closely, while ReLU, Swish, and Softmax show slightly higher errors and lower R^2 values. Overall, Tanh outperforms the other activation functions in this testing phase, demonstrating superior predictive accuracy and model fit.

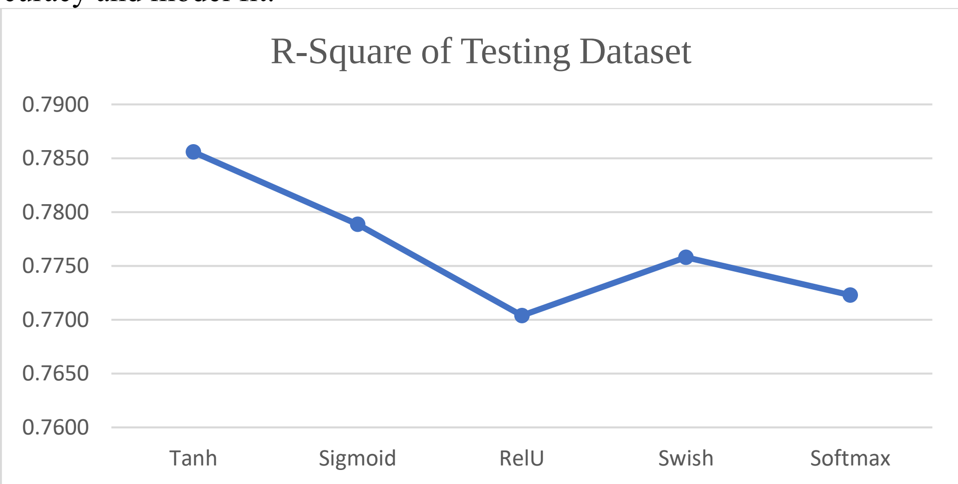

Figure 2. R-Square (R^2) values

The figure 1. presents the R-Square (R^2) values for different activation functions in an LSTM (Long Short-Term Memory) model applied to a testing dataset. Tanh has the highest R^2 value (0.78557), indicating the best predictive performance among the tested functions. Sigmoid follows closely with 0.77886, showing slightly lower accuracy. ReLU exhibits the lowest R^2 (0.77038), suggesting it is the least effective for this dataset. Swish improves over ReLU but does not outperform Tanh or Sigmoid, with an R^2 of 0.77580. Softmax also performs better than ReLU but falls slightly below Swish, with an R^2 of 0.77228. The overall trend suggests that Tanh is the most effective activation function for this LSTM model, while ReLU is the least suitable in this case.

Table (7): represents the forecasted values for 30 Days

Forecasted	Confidence Interval 95%	
	Lower	Upper
610267.1728	601152.127	618763.219
639518.5215	634244.02	644869.968
675384.3956	668724.065	681871.033
464680.2797	456742.843	472212.202
656281.2388	647892.124	664182.658
608629.2507	600372.204	616422.544
955189.2707	944949.441	964606.191
666467.2428	655808.641	676226.719
552822.987	545076.058	560198.938
611980.5195	604260.774	619334.214
617412.5974	606034.002	627761.541
580387.3327	571270.967	588884.754
656993.0723	639910.19	670170.229
648464.9743	631466.529	714548.679
703917.5281	686370.048	792416.207
726254.9494	708486.306	767377.646
736567.6439	718696.895	745440.234
657182.2889	640097.534	726672.33
753549.7205	735510.832	758212.325
648983.3284	631979.751	715101.663
891096.6082	871695.869	1055138.08
600232.7668	583711.868	646542.48
694574.9666	677119.987	704248.698
721626.1892	703903.375	731232.206
759479.8061	741382.204	777352.031
689773.3619	672365.923	715804.588
693424.7631	675981.171	720902.876
700807.0133	683290.33	731230.249
625406.9567	608636.809	681697.221
713129.364	695490.677	718107.302

The provided table shows forecasted values along with their 95% confidence intervals for each forecast. For each forecasted value, a lower and upper bound is given, which indicates the range within which the true value is likely to fall with 95% certainty. For example, the first forecasted value is 610,267.17, and its 95% confidence interval ranges from 601,152.13 to 618,763.22. This means there is a 95% probability that the actual value will fall within this interval. Similarly, the confidence intervals for all the forecasted values give a measure of the uncertainty in the predictions, with each interval showing the range of likely outcomes. The intervals vary in size depending on the forecast, indicating the level of confidence in the predictions and their precision.

Conclusion:

The study demonstrates that activation function choice significantly impacts LSTM model performance in time series forecasting. Among the tested functions, Tanh, Sigmoid, and ReLU exhibited varying degrees of accuracy, with the Tanh-activated LSTM achieving the highest R^2 value of 0.7837 at epoch 45, closely followed by the Sigmoid-activated LSTM at epoch 50 with an R^2 of 0.7794. The ReLU function also provided competitive results but showed fluctuations in accuracy across epochs. These findings suggest that careful selection of activation functions is crucial for optimizing LSTM models, with Tanh and Sigmoid performing better in stabilizing and improving predictive accuracy. Future work may explore hybrid activation functions or advanced optimization techniques to enhance forecasting precision further.

Reference:

1- Ahmed, B. K., Rahim, S. A., Maaroof, B. B., & Taher, H. A. (2020). Comparison Between ARIMA And Fourier ARIMA Model To Forecast The Demand Of Electricity In Sulaimani Governorate. *Qalaai Zanist Journal*, *5*(3), 908-940.

2- Ahmed, S., Mubarak, S., Du, J. T., & Wibowo, S. (2022). Forecasting the status of municipal waste in smart bins using deep learning. *International Journal of Environmental Research and Public Health*, *19*(24), 16798.

3- Asraa, A., Rodeen, W., & Tahir, H. (2018). Forecasting the Impact of Waste on Environmental Pollution. *International Journal of Sustainable Development and Science*, *1*(1), 1-12.

4- Aziz, A. A., Shafeeq, B. M., Ahmed, R. A., & Taher, H. A. (2023). Employing Recurrent Neural Networks to Forecast the Dollar Exchange Rate in the Parallel Market of Iraq. *Tikrit Journal of Administrative and Economic Sciences*, *19*(62), 2.

5- Bansal, M., Goyal, A., & Choudhary, A. (2022). A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short-term memory algorithms in machine learning. *Decision Analytics Journal*, *3*, 100071.

6- Cubillos M. (2020). Multi-site household waste generation forecasting using a deep learning Waste Management, https://doi.org/10.1016/j.wasman.2020.06.046.

7- da Silva, D. G., & de Moura Meneses, A. A. (2023). Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction. *Energy Reports*, *10*, 3315-3334.

8- Ding, Z., Yu, Y., & Xia, Y. (2023). Nonlinear hysteretic parameter identification using an attention-based long short-term memory network and principal component analysis. *Nonlinear Dynamics*, *111*(5), 4559-4576.

9- Fatovatikhah, F., Ahmedy, I., & Noor, R. M. (2024). Waste Prediction Approach Using Hybrid Long Short-Term Memory with Support Vector Machine. *International Journal of Computational Intelligence Systems*, *17*(1), 103.

10-Fu H, Li Z, Wang R. (2015). Estimating municipal solid waste generation by different activities and various resident groups in five provinces of China. Waste Management, 41, 3–11. https://doi.org/10.1016/j.wasman.2015.03.029.

11-Ghanbari F, Kamalan H, Sarraf A. (2021). An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. Arab J Geoscience, 14(2), 1-16. https://doi.org/10.1007/s12517-020-06348-w.

12-Huang L, Cai T, Zhu Y, Zhu Y, Wang W, Sun K. (2020). LSTM-Based Forecasting for Urban Construction Waste Generation. Sustainability, 12 (20), 8555. https://doi.org/10.3390/su12208555.

13-Karim, A. J. M., & Ahmed, N. M. (2023). Vector Autoregressive Integrating Moving Average (Varima) Model of COVID-19 Pandemic and Oil Price. *International Journal of Professional Business Review: Int. J. Prof. Bus. Rev.*, *8*(1), 13.

14-Mercioni, M. A., & Holban, S. (2020, May). The most used activation functions: Classic versus current. In *2020 International Conference on Development and Application Systems (DAS)* (pp. 141-145). IEEE.

15-Sha, X. (2024). Time Series Stock Price Forecasting Based on Genetic Algorithm (GA)-Long Short-Term Memory Network (LSTM) Optimization. *arXiv preprint arXiv:2405.03151*.

16-Taher, H. A., Salih, K. K., & Alwan, A. S. (2024). Forecasting For Silver Closing Price and Modifying Predictions by Using GARCH (1, 1) Wavelet Transformation. *University of Kirkuk Journal For Administrative and Economic Science*, *14*(2).

17-Utku, A., & Kaya, S. K. (2022). Deep learning based on a comprehensive analysis for waste prediction. *Operational Research in Engineering Sciences: Theory and Applications*, *5*(2), 176-189.

18-Yazdani, M., Kabirifar, K., & Haghani, M. (2024). Optimizing post-disaster waste collection by a deep learning-enhanced differential evolution approach. *Engineering Applications of Artificial Intelligence*, *132*, 107932.

19-https://images.app.goo.gl/QiZ2cYj6MfrZqCbGA.