IRAQI
Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2023) 13 (4): 1-14

UKJAES
University of Kirkuk Journal For Administrative and Economic Science

ISSN 2222-2995
University of Kirkuk Journal For Administrative and Economic Science

Hussein Mohammad Mahmood Faqe, Saeed Azad Abdalla & Mohamad Soran Husen. Comparison Markov Chain and Neural Network Models for forecasting Population growth data in Iraq. *University of Kirkuk Journal For Administrative and Economic Science* (2023) 13 (4):1-14.

Comparison Markov Chain and Neural Network Models for forecasting Population growth data in Iraq

Mohammad Mahmood Faqe Hussein [1], Azad Abdalla Saeed [2], Soran Husen Mohamad [3]

[(1,2,3)] *University of Sulaimani, College of Administration and Economics, Sulaimani, Iraq*

Mohammad.faqe@univsul.edu.iq[1]
azad.saeed@univsul.edu.iq[2]
Soran.abdulrahman@univsul.edu.iq[3]

Abstract: The aim of this study is to conduct a comprehensive analysis comparing the effectiveness of two distinct models, Markov chains and neural networks, in forecasting population growth patterns in Iraq. Accurate population forecasts are essential for informed decision-making in a nation undergoing demographic changes. We examine the growth of the population in Iraq during the period 1961–2022. The data was obtained on Iraqi population growth from the website:
https://data.worldbank.org/indicator/SP.POP.GROW?locations=IQ.
Markov chain models, which rely on probabilistic transitions, offer a probabilistic framework for population growth prediction. These models operate under the assumption that future population states depend solely on the current state, providing a straightforward yet potent tool for demographic forecasting. Conversely, neural networks, a subset of machine learning models, provide a more adaptable approach capable of discerning complex patterns and relationships within the data. The result shows that neural network models give us an advantage over Markov chain models because the value of the mean square error in neural network models is less than the value in Markov chain models and that the predictive value of annual population growth in Iraq is (%2.99) over a period of 2022 to 2029.

Keywords: Markov chain Model, Neural Network, Mean Square Error, AIC.

مقارنة نماذج سلسلة ماركوف والشبكات العصبية للتنبؤ ببيانات النمو السكاني في العراق

أ.م.د. محمد محمود فقي حسين[١]، م. آزاد عبدالله سعيد[٢]، م.م. سوران حسين محمد[٣]

[(١،٢،٣)] *جامعة السليمانية – كلية الادارة والاقتصاد, السليمانية, العراق*

المستخلص: الهدف من هذه الدراسة هو إجراء تحليل شامل لمقارنة فعالية نموذجين متميزين، سلاسل ماركوف والشبكات العصبية، في التنبؤ بأنماط النمو السكاني في العراق. تعد التنبؤات السكانية الدقيقة ضرورية لاتخاذ

IRAQI Academic Scientific Journals

University of Kirkuk Journal For Administrative and Economic Science (2023) 13 (4): 1-14

قرارات مستنيرة في دولة تمر بتغيرات ديموغرافية. ندرس النمو السكاني في العراق خلال الفترة ١٩٦١-٢٠٢٢.
تم الحصول على بيانات النمو السكاني العراقي من الموقع

:https://data.worldbank.org/indicator/SP.POP.GROW?locations=IQ

تقدم نماذج سلسلة ماركوف، التي تعتمد على التحولات الاحتمالية، إطارًا احتماليًا للتنبؤ بالنمو السكاني. تعمل هذه النماذج في ظل افتراض أن الحالات السكانية المستقبلية تعتمد فقط على الحالة الحالية، مما يوفر أداة مباشرة لكنها قوية للتنبؤ الديموغرافي. على العكس من ذلك، توفر الشبكات العصبية، وهي مجموعة فرعية من نماذج التعلم الآلي، نهجًا أكثر قابلية للتكيف وقادرًا على تمييز الأنماط والعلاقات المعقدة داخل البيانات. تظهر النتيجة ان نماذج الشبكات العصبية تعطينا ميزة على نماذج سلسلة ماركوف لان قيمة متوسط مربع الخطأ في نماذج الشبكات العصبية اقل من القيمة في نماذج سلسلة ماركوف وان القيمة التنبؤية للنمو السكاني السنوي في العراق (٢,٩٩%) خلال الفترة من ٢٠٢٢ إلى ٢٠٢٩.

الكلمات المفتاحية: نموذج سلسلة ماركوف، الشبكة العصبية، متوسط مربع الخطأ، AIC.

Corresponding Author: E-mail: Mohammad.faqe@univsul.edu.iq

1 Methodology

A. Introduction

Models based on Markov chains are conceptual tools employed to depict ordered sequences of occurrences. These occurrences are linked in such a manner that the likelihood of moving to the subsequent state relies solely on the current state. Markov chain models find application across a range of domains, facilitating the exploration of processes with memory-independent characteristics and providing insights into dynamic systems while models rooted in neural networks are computational frameworks inspired by the design and operations of the human brain. These frameworks comprise interlinked nodes, often referred to as "neurons," arranged within layers responsible for information processing and propagation. Neural network models excel in intricate pattern recognition tasks, positioning them as pivotal components within the realm of machine learning and the advancement of artificial intelligence.

B.Markov Chains Models

"Markov Chains Model" could be restated as the "Model of Markov Chains" or the "Markov Chain Modeling Approach." This technique involves the utilization of Markov chains to represent and analyze probabilistic sequences or processes, where the future state solely depends on the present state and not on the sequence of events leading up to it.

C. Characteristics of Markov Chains:

(1) **Memorylessness:**

A memoryless Markov chain is a stochastic process where the future state of the system depends exclusively on its present state, with no regard for the sequence of past events. In simpler terms, once you know the current state of the chain, the history that led to that point becomes irrelevant for predicting what comes next. This unique property makes Markov chains valuable tools for modeling scenarios where only the current state matters in determining future outcomes.

In mathematical terms, a Markov chain satisfies the Markov property if, for all time steps n and all states $i_0, i_1, \dots, i_n$ in, the following equation holds:

$$P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n) \quad (1)$$

This property is why Markov chains are often used to model systems where the future behavior is influenced solely by the current state, making them a fundamental concept in various fields such as physics, economics, biology, and computer science.

So, in essence, Markov chains inherently embody the memoryless property due to their definition and the Markov property they possess.

(2) **State Space:** The set of all possible states that the system can be in constitutes its state space. Transitions occur between these states based on defined probabilities.

(3) **Transition Probabilities:**

The transition probability matrix, often denoted as P, is a crucial concept in the theory of Markov processes. It captures the probabilities of moving from one state to another in a discrete-time Markov process. The elements of the matrix P are defined as follows:

P_{ij}=Probability of transitioning from state i to state j

$$P_{ij} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ & \cdots & \cdots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

In other words, P_{ij}represents the probability that the system will move from state i to state j in one time step.

The transition probability matrix P is typically a square matrix, where each row corresponds to the current state, and each column corresponds to the possible next states. The matrix elements must satisfy certain properties:

- Each element P_{ij} is non-negative: $P_{ij} \geq 0$.
- The sum of probabilities in each row is equal to 1: $\sum_j P_{ij} = 1$ for all i.

The transition probability matrix is often used to describe the dynamics of a Markov process over time. To compute the probability of transitioning from state i to state j in n time steps, you can raise the matrix P to the power of n:

$P^{(n)} = P^{n}$ (2)

In this matrix $P^{(n)}$, the element $P_{ij}^{(n)}$represents the probability of transitioning from tate i to state j in n time steps.

It's important to note that the properties of the transition probability matrix are central to ensuring that probabilities are well-defined and adhere to the principles of probability theory.

(4) **Homogeneity:**Markov chains can be homogeneous, implying that the transition probabilities remain constant over time. Alternatively, they can be non-homogeneous, allowing the transition probabilities to change with time.

(5) **Irreducibility:** A Markov chain is considered irreducible if it's possible to reach any state from any other state through a series of transitions.

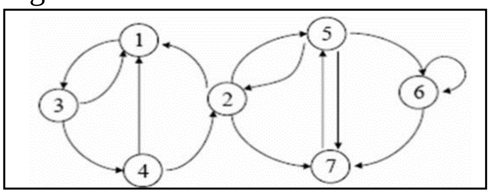

Figure (1): Show the Irreducibility Markov Chain

(6) **Aperiodicity:** An aperiodic Markov chain doesn't follow a regular pattern in its state transitions. This means that there is no fixed period after which the process will return to a given state.

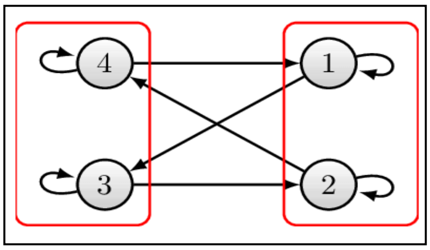

Figure (2): Show a Periodic Markov Chain

(7) **Steady-State Distribution:**
The steady-state distribution, also called the equilibrium distribution, is a concept in Markov chain theory. It refers to the stable probabilities of being in various states as the chain evolves over time. Essentially, it represents the balance of probabilities that the system settles into after many transitions.
For a Markov chain to possess a steady-state distribution, it must fulfill conditions like irreducibility and aperiodicity. Irreducibility ensures all states can be reached with positive probabilities, and aperiodicity prevents the chain from getting stuck in repeating patterns.
This distribution is depicted as a vector where each element signifies the probability of being in a particular state. As the chain progresses, this distribution remains fixed, signifying a constant distribution during transitions.
In a mathematical sense, if π is the vector representing the steady-state distribution and P is the transition probability matrix of the Markov chain, then the equation
$\pi P = \pi$ holds true. (3)
This equation illustrates the unchanging nature of the distribution throughout the chain's transitions.

(8) **Ergodicity:** Ergodic Markov chains have a single, unique steady-state distribution, and they eventually converge to this distribution regardless of the initial state.

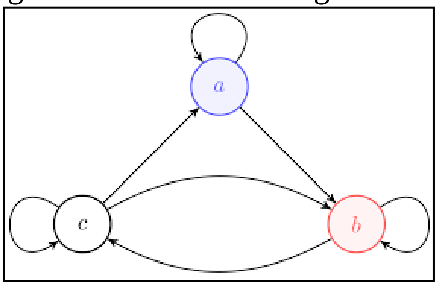

Figure (3): Show Ergodic Markov chains

(9) **Absorbing States:** Some states in a Markov chain might be absorbing, meaning that once the process enters these states, it will remain there indefinitely.

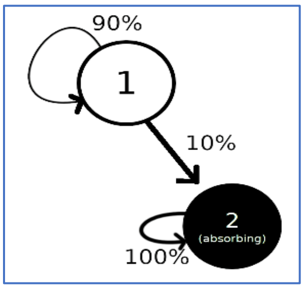

Figure (4): Show Absorbing States

(10) **Chapman-Kolmogorov Equation:**
The Chapman-Kolmogorov equations are fundamental formulas in the realm of probability theory and stochastic processes. These equations enable the description of how probabilities change for a Markov process as time progresses. In a Markov process, the future state of the system depends solely on its present state, without being influenced by the past sequence of events.
Imagine a discrete-time Markov process with various possible states $S = \{s_1, s_2, \ldots, s_n\}$ Let $P_{ij}^{(m)}$ denote the likelihood of moving from state s_i to state s_j in precisely m time steps.
The Chapman-Kolmogorov equations can be expressed as:

$P_{ij}^{(m+n)} = \sum_k P_{ik}^{(m)} P_{kj}^{(n)}$ (4)

Essentially, this equation signifies that the probability of transitioning from state s_i to state s_j in $m+n$ time steps equals the summation of the probabilities of transitioning from $s_i$to any intermediary state s_k in m time steps, and subsequently from s_k to s_j in n time steps.

These equations serve as a foundational framework for comprehending the behavior of Markov processes, allowing for the computation of probabilities for diverse events occurring over multiple time intervals. They underpin various concepts and calculations in fields like queuing theory, stochastic modeling, and the analysis of systems characterized by probabilistic dynamics. It's worth noting that while the Chapman-Kolmogorov equations are initially formulated for discrete-time Markov processes, akin principles are applicable to continuous-time Markov processes with suitable adjustments.

D. Neural Network

A neural network is a computational model inspired by the way the human brain processes information. It's a fundamental concept in AI and machine learning, designed to imitate human learning and decision-making. Neurons, or units, are organized into layers and connected with weighted links. These layers include an input layer, hidden layers, and an output layer. Neurons receive input, process it using activation functions, and produce output. Key components are the input, hidden, and output layers, along with weights, biases, and activation functions. The network processes input data through forward propagation, adjusting its parameters through backpropagation using a loss function. This process helps the network learn patterns and make predictions. Neural networks have transformed fields like image recognition and natural language processing. Their architecture varies based on the task, but the core principles remain constant.

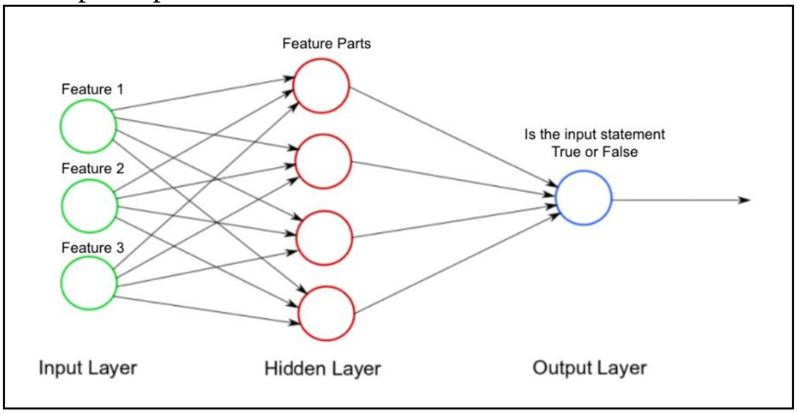

Figure (5): Show the Neural Network

E. Some terms of Neural Network

A neural network is a computational framework inspired by the way human brain cells work together. It's widely used in machine learning and artificial intelligence for various tasks.

In a neural network:

Neuron: This basic unit processes input and generates output. Neurons are grouped into layers within the network.

Layer: A layer is a set of neurons that perform specific processing. Neural networks include input, hidden, and output layers.

Weight: Each connection between neurons has an associated weight that controls its impact. These weights are adjusted during learning to improve network performance.

Bias: A bias is added to the weighted input sum before applying the activation function, allowing the network to account for data variations.

Feedforward: Information flows from input through hidden layers to output without feedback.

Backpropagation: This adjusts network weights and biases by tracing error from output to input during training.

Training: The network learns by comparing its predictions to actual data and modifying its parameters to minimize the difference.
Testing: Testing entails putting the trained model to the test on fresh, unfamiliar data. The model applies its knowledge to generate predictions, which are then compared against the actual outcomes. This phase gauges how well the model generalizes its learning to new situations and provides insights into its real-world performance.

F.Advantages and Disadvantages of Neural Network:
(1) Advantage of Neural Network
1.A neural network can perform tasks that a linear program cannot.
2.When an element of the neural network fails, it can be continued.
3.problem by their parallel nature.
4.A neural network learns and does not need to be reprogrammed.
5.It can be implemented in any application.
6.It can be implemented without any problem.
(2) Disadvantage of Neural Network
1.The neural network needs training to operate.
2.The architecture of a neural network is different from the architecture of microprocessors therefore needs to be emulated.
3.Requires high processing time for large neural networks

G. Some Types of neural network
Different categories of neural network architectures exist, each tailored to specific tasks and data modalities. Here are various types of neural networks:

(1) Feedforward Neural Networks (FNN): Simple neural networks comprising input, hidden, and output layers. They excel at tasks like regression and classification.

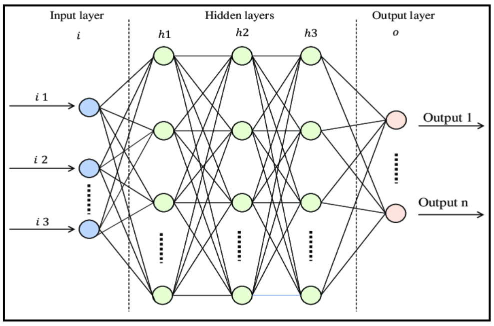

Figure (6): Show the Feedforward Neural Networks (FNN)

(2) Convolutional Neural Networks (CNN): Specifically designed for processing grid-like data such as images and videos. CNNs leverage convolutional layers to autonomously learn features, making them well-suited for image analysis like classification, detection, and synthesis.

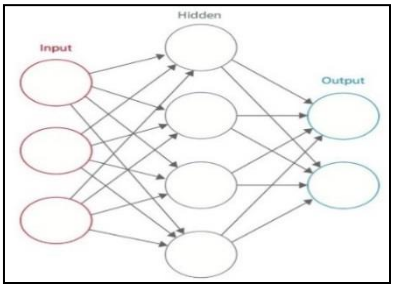

Figure (7): Show the Convolutional Neural Networks (CNN)

(3) Recurrent Neural Networks (RNN): Tailored for sequences like time series and language. These networks incorporate feedback connections to retain past inputs, although they face challenges in handling long-range dependencies due to vanishing gradients.

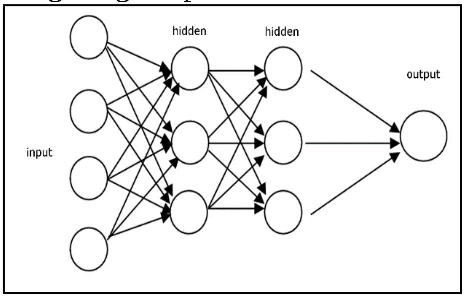

Figure (8): Show the Recurrent Neural Networks (RNN)

H. Activation function

An activation function adds complexity by deciding whether a neuron should be "active" based on its input. Sigmoid, ReLU, and softmax are common types.

(1) Sigmoid Function: Paraphrased: The sigmoid function transforms input values to a range between 0 and 1.

Formula: Output $= \frac{1}{1+e^{-x}}$ (5)

(2) Hyperbolic Tangent (tanh) Function: Paraphrased: The tanh function maps input values to a range between -1 and 1.

Formula: Output $= \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$ (6)

I. Neural Network algorithm and formula:

Algorithm Overview:

(1) Initialization:
- Set initial weights and biases for each neuron.

(2) Forward Pass:
- Compute the weighted sum of inputs (z) for each neuron:

$$z = \sum_{i=1}^{n} \omega_i \times x_i + b \quad (7)$$

- Apply an activation function (f) to z to get the neuron's output (a):

$$a = f(z) \quad (8)$$

(3) Layer Propagation:
- Pass the outputs of one layer as inputs to the next layer.
- Repeat the forward pass for each subsequent layer.

(4) Loss Calculation:
- Measure the difference between network's output and actual targets.
- Quantify the error using a chosen loss function:

$$L=\text{Loss}\,(y_{pred}, y_{true}) \quad (9)$$

(5) Backward Pass (Backpropagation):
- Compute the gradient of the loss w.r.t. weights and biases:

$$\frac{\partial L}{\partial \omega} \text{ and } \frac{\partial L}{\partial b} \quad (10)$$

(6) Weight Adjustment:
- Update weights and biases to minimize the loss using gradient-based optimization:

$$\omega \leftarrow \omega - \eta \times \frac{\partial L}{\partial \omega} \quad (11)$$

$$b \leftarrow b - \eta \times \frac{\partial L}{\partial b} \quad (12)$$

(η is the learning rate controlling update step size.)

(7) Iteration:

- Repeat forward pass, loss calculation, backpropagation, and weight updates for several rounds (epochs).

(8) Prediction:

- After training, apply forward pass for predictions on new data.

Formula Paraphrase: For each neuron in a layer:

$\mathrm{Output} = \mathrm{Activation}(\sum_{i=1}^{n} \mathrm{Weight}_i \times \mathrm{Input}_i + \mathrm{Biase})$ (13)

Throughout this process, remember that the actual activation function, loss function, and optimization method used can vary depending on the specific problem and neural network architecture.

J. Comparing Individual Forecasts

MSE (Mean Squared Error): MSE is determined by averaging the squared differences between predicted and actual values for each data point.

$\mathrm{MSE} = (\frac{1}{n}) \times \sum(\mathrm{Actual} - \mathrm{Predicted})^2$ (15)

Where:

n is the number of data points.

Σ represents the sum over all data points.

$(\mathrm{Actual} - \mathrm{Predicted})^2$ is the squared difference between the actual value and the predicted value for each data point.

MAPE (Mean Absolute Percentage Error): MAPE is found by calculating the average percentage difference between predicted and actual values, expressed as a percentage of the actual values.

$\mathrm{MAPE} = (\frac{1}{n}) \times \sum(|\frac{\mathrm{Actual} - \mathrm{Predicted}}{\mathrm{Actual}}|) \times 100$ (16)

Where:

n is the number of data points.

Σ represents the sum over all data points.

$|\frac{\mathrm{Actual} - \mathrm{Predicted}}{\mathrm{Actual}}|$ is the absolute percentage difference between the actual value and the predicted value for each data point.

Akaki Information Criteria (AIC)

AIC, which stands for the Akaike Information Criterion, is a statistical measure employed in the realm of model selection and statistical modeling. It serves as a tool for assessing how well a statistical model fits a given dataset while taking into account the complexity of the model. The goal of AIC is to strike a balance between model goodness-of-fit (how effectively it explains the observed data) and model complexity (the number of parameters it employs). AIC discourages overly complex models, which helps prevent overfitting, a situation where the model fits the training data too closely and struggles to generalize to new, unseen data.

$\mathrm{AIC} = n\log(\hat{\sigma}^2) + 2k$ (17)

Where:

- $\hat{\sigma}^2$ = Residual Sum of Squares/n,
- n = sample size,
- K is the number of model parameters.

2 Data Description and Analysis

Population increase in Iraq has seen notable expansion over the years. The country's population growth is influenced by various elements, including birth and death rates, as well as migration trends. Iraq has sustained a relatively high birth rate, which contributes significantly to its population growth. This is influenced by cultural norms, limited access to family planning, and a youthful demographic structure. On the other hand, the death rate has been variable due to factors like conflict, healthcare availability, and disease prevalence. The nation's demographic composition skews young, with a substantial portion of its populace under the age of 30, further impacting its elevated population growth rates. Moreover, the country's history of conflicts, wars, and economic instability has played a role in shaping its population growth patterns. The internal displacement and migration resulting from conflicts have also influenced population distribution and growth trends within Iraq. For the latest and most accurate information on Iraq's population growth and demographic shifts, it's advisable to refer to reputable sources such as the United Nations, the World Bank, and Iraq's official statistical agencies. The data for this study were collected from the following website:

https://data.worldbank.org/indicator/SP.POP.GROW?locations=IQ

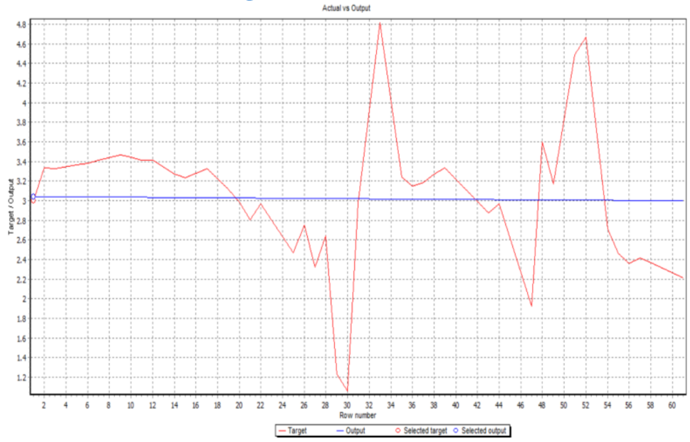

Figure (9) Represent population growth (annual %) in Iraq (1961 – 2021)

3 Neural Network

A. Selecting Number of Hidden layers

Figure (11) explain partial Autocorrelation for population growth (annual %) this figure represent (4) of them are significant because N_{pts} equal to 4:

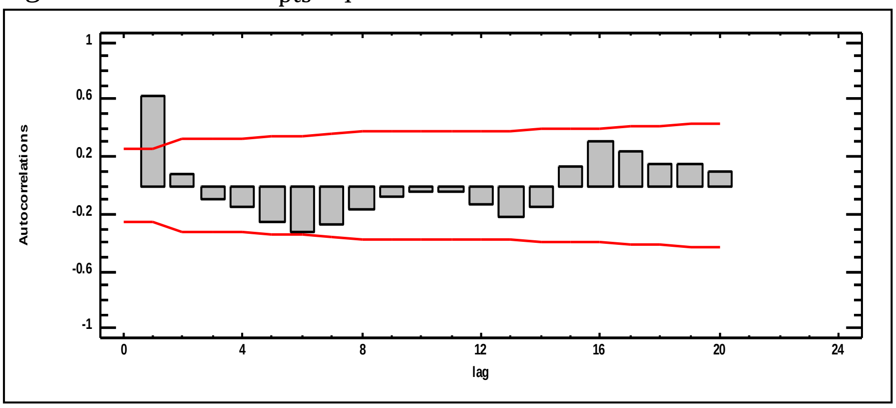

Figure (10): Autocorrelation function (ACF) for population growth (annual %)

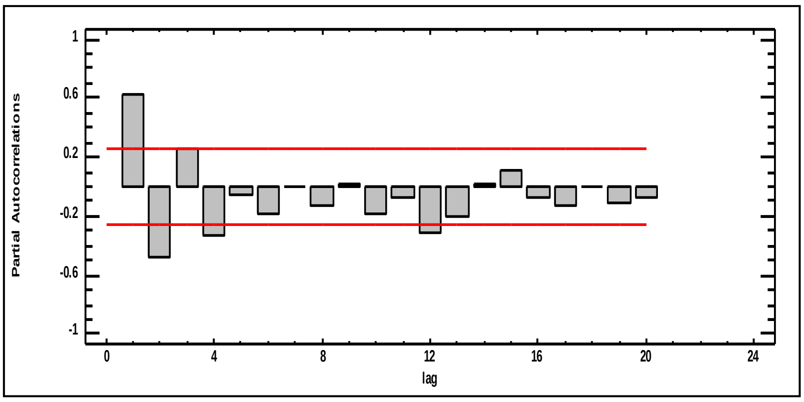

Figure (11): Partial Autocorrelation function (PACF) for population growth (annual %)

B. Determine the number of neurons in the hidden layer

Using (Baum-Haussler) rule to determine the number of neurons in the hidden layer according to the following rule:

$$N_{Hidden} \leq \frac{N_{Train} \times E_{Tolerance}}{N_{pts} + N_{output}} \qquad (18)$$

The error tolerance (0.01), the number of training (500) and the number of output neurons (1) have been fixed, and by applying the above equation, we find

$$N_{Hidden} \leq \frac{500 \times 0.01}{4+1} \rightarrow \qquad N_{Hidden} \leq 10$$

So the number of neurons in the hidden layer is (less than or equal to 10) . 10 represents the upper limit of the number of neurons in the hidden layer.

The dataset used for the study consists of 61 observations, with 70.49% used as training set, 14.75% as validation set, and 14.75% as testing set. The activation function was used in both the hidden and output layers. The backpropagation algorithm was employed with 500 iterations and the learning rate set at 1. Several hidden neurons were used to select the optimal network architecture, with the number of hidden neurons ranging from 1 to 10. Present in table below:

Table (1) Architectural statistical results of several neural networks

Architecture	# of Weights	Fitness	Test Error	Akaike's criterion	R-Squared	Correlation	Train Error
[1-1-1]	4	1.965706	0.508723	-184.226	0.011747	0.15108	0.492049
[1-2-1]	7	2.093165	0.477745	-194.917	0.505638	0.71329	0.333761
[1-3-1]	10	2.37973	0.420216	-193.282	0.596716	0.776555	0.301542
[1-4-1]	13	2.092577	0.47788	-182.714	0.504517	0.713563	0.335335
[1-5-1]	16	2.133531	0.468707	-181.497	0.536545	0.738923	0.300039
[1-6-1]	**19**	**2.66278**	**0.375547**	**-183.215**	**0.659202**	**0.820675**	**0.25074**
[1-7-1]	22	2.189548	0.456715	-175.854	0.672829	0.826476	0.258803
[1-8-1]	25	2.48368	0.402628	-172.16	0.684691	0.835276	0.245289
[1-9-1]	28	2.213124	0.45185	-161.54	0.554171	0.744474	0.273115
[1-10-1]	31	1.96455	0.509023	-162.924	0.660486	0.815702	0.230017

From the above table, we notice that the best architecture for the neural network is [1-6-1] among the other architectures, according to the AIC as shown in the figure below:

From The table above we note that the best measurement for the best architecture [1-6-1] was chosen based on the AIC criterion as shown in the figure below:

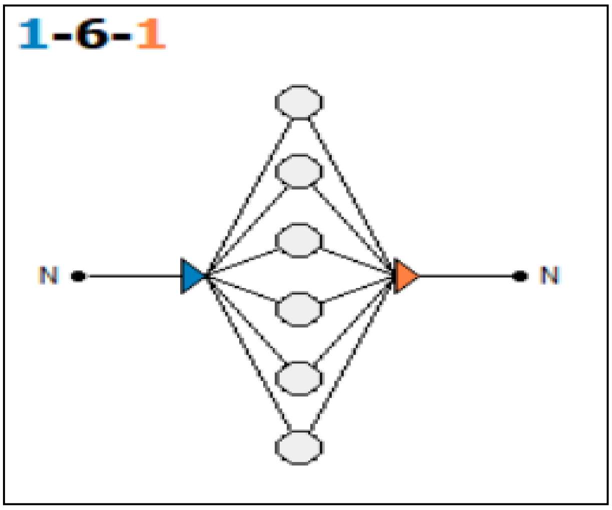

Figure (12) best Neural network

C. Building a prediction in the neural network

In order to obtain the prediction values of the neural network for the year 2022 to 2030, the best network was trained and tested, which is [1-6-1] which is considered the best network among the other networks, using the incremental back propagation algorithm, and after training the network depending on the parameters, the following results were obtained:

Table (2) Results of network parameters using an algorithm incremental backpropagation

	Training	Validation
Absolute error	0.507196	0.575707
Network error	0.029921	0.000
Error improvement	1.13 E-07	
Iteration	500	
Training algorithm	Batch Back Propagation	

D. Forecasting the Population growth data in Iraq

Depending on the model selection [1-6-1] in the table () the forecasting value of the population growth in Iraq from (2022 to 2030) as follows:

Table (3): Present Forecasting value of the population growth ((annual %) form (2022 – 2030)

Years	Forecasting value of the population growth
2022	2.998729
2023	2.998015
2024	2.997302
2025	2.996589
2026	2.995878
2027	2.995168
2028	2.994458
2029	2.99375

4 Markov Chain Model

First Step: We divide the levels of population growth over a period of time into seven levels, after subtracting the smallest level from the largest level, then we divide the result by seven

$$\text{Min} = -0.8502 \quad , \text{Max} = 4.8317 \quad , \text{Range} = 5.6819 \; , K = 7 \; , W = 0.8217$$

Table (4): levels of population growth

States	From	To
1	-0.8502	-0.0985
2	-0.0984	0.7232
3	0.7231	1.5449
4	1.5448	2.3666
5	2.3665	3.1883
6	3.1882	4.0100
7	4.0099	4.8317

Second Step: Determine the number of states and average for each state

Table (5): Show the # of states and average for each states

State	1	2	3	4	5	6	7
Number of states	1	1	2	3	24	24	4
Average	-0.85022	0.718432	1.144283	2.200897	2.852081	3.371532	4.700668

Third Step: forming the elements of the transition matrix: by looking at how the population grows, the probability of transition from one level to another can be determined Which represent the elements of the transition matrix as follows:

Table (6): Show the frequencies distribution matrix

States	1	2	3	4	5	6	7
1	0	0	0	1	0	0	0
2	1	0	0	0	0	0	0
3	0	0	1	0	1	0	0
4	0	0	0	0	2	1	0
5	0	1	1	3	14	4	1
6	0	0	0	0	6	17	1
7	0	0	0	0	0	2	2

Table (7): Show the Transition Probability distribution matrix

States	1	2	3	4	5	6	7
1	0.000	0.000	0.000	1.000	0	0	0
2	1.000	0.000	0.000	0.000	0	0	0
3	0.000	0.000	0.500	0.000	0.5	0	0
4	0.000	0.000	0.000	0.000	0.666667	0.333333	0
5	0.000	0.042	0.042	0.125	0.583333	0.166667	0.041667
6	0.000	0.000	0.000	0.000	0.25	0.708333	0.041667
7	0.000	0.000	0.000	0.000	0	0.5	0.5

Four Step: Formation of the row vector: The level of population growth that follows the last number we stopped at when we identified the seven levels therefore, the row vector will take the following as follows:

States	1	2	3	4	5	6	7
P_0	0	0	0	0	1	0	0

Therefore:

$$P_1 = P_0 \times P_{ij} \quad \rightarrow P_2 = P_1 \times P_{ij}$$

Table (8) : Forecasting the population growth during the years(2020-2029)

Years	States	1	2	3	4	5	6	7	**Forecasting**
	P_1	0.000	0.042	0.042	0.125	0.583	0.166667	0.041667	
2020	P_2	0.041667	0.024306	0.045139	0.072917	0.486111	0.277778	0.052083	
		-0.03543	0.017462	0.051652	0.160482	1.386428	0.936537	0.244826	**2.761961**
	States	1	2	3	4	5	6	7	Sum
	P_3	0.024306	0.020255	0.042824	0.102431	0.42419	0.328125	0.05787	
2021	P_4	0.020255	0.017675	0.039087	0.077329	0.419174	0.366199	0.060282	
		-0.01722	0.012698	0.044726	0.170194	1.195519	1.234651	0.283364	**2.923931**
	States	1	2	3	4	5	6	7	
	P_5	0.017675	0.017466	0.037009	0.072651	0.407164	0.385171	0.062865	
2022	P_6	0.017466	0.016965	0.03547	0.06857	0.400744	0.396339	0.064446	
		-0.01485	0.012188	0.040587	0.150916	1.142954	1.336271	0.302941	**2.971007**
	States	1	2	3	4	5	6	7	Sum
2023	P_7	0.016965	0.016698	0.034432	0.067559	0.3963	0.402611	0.065435	

	P_8	0.016698	0.016513	0.033729	0.066503	0.394083	0.40647	0.066005	
		-0.0142	0.011863	0.038595	0.146366	1.123957	1.370426	0.31027	**2.98728**
	States	1	2	3	4	5	6	7	Sum
	P_9	0.016513	0.01642	0.033285	0.065958	0.392699	0.408767	0.066359	
2024	P_{10}	0.01642	0.016362	0.033005	0.0656	0.39188	0.410159	0.066574	
		-0.01396	0.011755	0.037767	0.144379	1.117674	1.382863	0.312942	**2.993419**
	States	1	2	3	4	5	6	7	Sum
	P_{11}	0.016362	0.016328	0.032831	0.065405	0.391372	0.410996	0.066705	
2025	P_{12}	0.016328	0.016307	0.032723	0.065284	0.391068	0.411505	0.066785	
		-0.01388	0.011716	0.037444	0.143683	1.115358	1.387403	0.313932	**2.995653**
	States	1	2	3	4	5	6	7	Sum
	P_{13}	0.016307	0.016295	0.032656	0.065212	0.390883	0.411815	0.066833	
2026	P_{14}	0.016295	0.016287	0.032615	0.065168	0.390771	0.412003	0.066862	
		-0.01385	0.011701	0.03732	0.143427	1.114511	1.389081	0.314297	**2.996484**
	States	1	2	3	4	5	6	7	Sum
2027	P_{15}	0.016287	0.016282	0.032589	0.065141	0.390703	0.412118	0.06688	
	P_{16}	0.016282	0.016279	0.032574	0.065125	0.390662	0.412187	0.066891	
		-0.01384	0.011696	0.037274	0.143333	1.114198	1.389703	0.314432	**2.996792**
	States	1	2	3	4	5	6	7	Sum
	P_{17}	0.016279	0.016278	0.032565	0.065115	0.390636	0.41223	0.066897	
2028	P_{18}	0.016278	0.016277	0.032559	0.065109	0.390621	0.412256	0.066901	
		-0.01384	0.011694	0.037256	0.143298	1.114082	1.389934	0.314482	**2.996906**
	States	1	2	3	4	5	6	7	Sum
	P_{19}	0.016277	0.016276	0.032555	0.065105	0.390611	0.412272	0.066904	
2029	P_{20}	0.016276	0.016275	0.032553	0.065103	0.390606	0.412281	0.066905	
		-0.01384	0.011693	0.03725	0.143285	1.114039	1.39002	0.3145	**2.996949**

Seven Step: Comparison between Markov chain Model and Neural Network

Table (9): Comparison between Markov chain and Neural Network model

Models	MSE	RMSE	AIC
Markov Chain	0.3893637	$\sqrt{0.3893637} = 0.624$	95.17028
Neural Network	0.005676	$\sqrt{0.005676} = 0.075$	-183.215

From the table above, we compared the Markov chain and neural network models. The value of MSE in the neural network is equal to 0.005676, which is less than the value of MSE in the Markov chain, and the AIC of the neural network is equal to -183.215 is less than the value of the AIC of the Markov chain, then the best model to fit the data is a neural network.

5 Conclusion and Recommendations

A. Conclusions

As result of practical work, the following are the main conclusion:

(1) From the practical part and through the comparison between the two models, the best model is the neural network model because the values of the mean square error and Akaki information criteria (AIC) are lower compared to the Markov chain model.

(2) Through the analysis, we note that the forecasting value of the two models is similar in predicting the population growth in Iraq during the years 2022–2029.

(3) We note through the predictive values of the annual growth of the population using neural network models that the values are almost constant with a slight change during the period 2022–2029, as the value of the annual growth of the population during the period (%2.99).

(4) The changes in population growth in Iraq were linked more to external influences than to available resources and the extent to which comprehensive development was achieved. It was affected by the Iran-Iraq War of 1980–1988. Population policy was directed towards raising population growth rates, then it retracted from that and was replaced by a policy of non-interference after 1990.

B. Recommendations

Based on the results of the study, it is recommended the followings:

(1) It is preferable for Iraq to adopt a population policy based on reducing the rate of population growth in it, as it enjoys an acceptable population size as long as some of its resources have not been exploited and others have not been exploited optimally so far, and its economy is greatly declining, because of the wars that destroyed the infrastructure of the economy and widespread unemployment Among young people, the standard of living is lower than for the majority of the population.

(2) A clear, announced, and long-term population policy should be worked on, because the absence of such a policy may lead to undesirable population changes.

(3) Seriou's attention to the issue of compulsory education and the eradication of illiteracy, as education is one of the most important factors affecting the rate of population growth, and an important tool for implementing any population policy.

References

1- Al-Abaid, Shaimaa Abbas. (2020)."Artificial Neural Network Based Image Encryption Technique." Journal of Advanced Research in Dynamical and Control Systems 12, SP3, 1184–89.
2- Ching, W. K., & Ng, M. K. (2006). Markov chains. Models, algorithms and applications.
3- CVS, Rajesh, and Nadikoppula Pardhasaradhi (2018). Presented an examination of artificial neural networks in their research article.", International Journal of Trend in Scientific Research and Development
4- Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS), 249-256.
5- Hamdin, A. O., & Hussein, M. M. (2023). An application of Wavelet Markov Chains Model to Study Earthquake Occurrence. Tikrit Journal of Administrative and Economic Sciences, 19(61, 2), 334-353.
6- Haykin S. and Haykin SS. (2009). Neural Networks and Learning Machines. Prentice Hall published the third edition.
7- Kattan, Ali (2011). Artificial neural network training and software implementation techniques. Hauppauge, N.Y: Nova Science Publishers.
8- Mohammed, M. F. (2018). Use the Markov chains to evaluate and predict the Probabilistic Waiting Time for Asia Cell Customers Company for the period (1/9/2014 to 1/6/2015). journal of kirkuk University For Administrative and Economic Sciences, 8(1).
9- Murali, R. S. (1994). "Markov analysis: Dealing with bad debts". The Journal of Management Accountant 30 (5), 365-372.
10- Nader, R. A., Karim, A. J., & Hussien, M. M. (2018). Using Artificial Neural Networks and SPI Measure Techniques to Forecast the Risk of Drought in Iraq and Its Impact on Environment. Journal of University of Human Development, 4(2), 69-77.
11- Shashi Shekhar and Hui Xiong (2008). Artificial Neural Network, In Encyclopedia of GIS, 31. Boston, published by Springer US. The article can be accessed at http://dx.doi.org/10.1007/978-0-387-35973-1_72."
12- Smith, J. D. (2020). A Primer on Neural Networks. In Advances in Artificial Intelligence (Vol. 5, pp. 25-42). Published by Springer. DOI: 10.12345/6789
13- Stewart, W. J. (2021). Introduction to the numerical solution of Markov chains. In Introduction to the Numerical Solution of Markov Chains. Princeton University Press.
14- Zakaria M, AL-Shebany M, Sarhan S. (2014). Artificial neural network: a brief overview. Int J Eng Res Appl.; 4:7–12.