AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

مجلة كلية الرافدين الجامعة للعلوم

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

استعمال تعليم الآلة في النماذج الإحصائية للتنبؤ السكاني في العراق

عمر عادل عبدالوهاب	مريم محمد سلمان
omaradil.d87@gmail.com	maryam90salman@gmail.com
قسم الاحصاء، كلية الادارة والاقتصاد، جامعة ديالى، ديالى، العراق	جامعة ابن سينا للعلوم الطبية والصيدلانية، بغداد، العراق

المستخلص

ان النماذج الاحصائية تعتبر الركيزة الاساسية لاي عملية تخطيط مستقبلي في اي بلد، اذ تستخدم الحكومات والقطاع الخاص التنبؤات السكانية في التخطيط. ان الغرض الأساسي للحكومة هو توفير الخدمات للمواطنين، وهذا يتطلب معرفة عدد الأشخاص الذين سيتواجدون في المستقبل، وغالبًا ما يتم تقسيمهم حسب العمر والجنس وخصائص أخرى، مثل العرق. والجغرافيا الى اخره. كما تُستخدم التنبؤات السكانية على نطاق واسع في القطاع الخاص للتخطيط الاستراتيجي، أذ تعتبر الاسقاطات السكانية من اهم المؤشرات الأدوات التي يستخدمها الباحثون وصانعو السياسات لفهم وتحليل تطورات السكان على مدى الوقت. تم في هذا البحث استخدام تعليم الآلة والمتمثلة بالشبكات العصبية، بالإضافة الى استخدام طريقة كلاسيكية وهي أنموذج السلاسل الزمنيةARIMA للتنبؤ بالاسقاطات السكانية. اذ تم تقسيم البيانات الداخلة في الاسلوبين الى 70% تدريب و30% اختبار، وتم التنبؤ لخمس سنوات قادمة لعدد سكان العراق. اذ توصل البحث الى افضلية تعليم الالة في التقدير كونه يمتلك اقل MAPE واعلىR2. حيث بلغ عدد سكان العراق لعام 2028 هو47516.3113 (بالالف).

معلومات البحث

تواريخ البحث:

تاريخ تقديم البحث: 18/2/2024
تاريخ قبول البحث: 12/4/2024
تاريخ رفع البحث على الموقع: 31/12/2024

الكلمات المفتاحية:

الاسقاطات السكانية، تعليم الآلة، الشبكات العصبية، ARIMA ، التنبؤ

للمراسلة

عمر عادل عبدالوهاب
omaradil.d87@gmail.com

DOI: https://doi.org/10.55562/jrucs.v56i1.8

1. المقدمة

الهدف الرئيسي للدولة هو تقديم خدمات عالية المستوى لمواطني البلاد. للقيام بذلك، من الضروري معرفة عدد الأشخاص الذين سيكونون في المستقبل، وفي بعض الحالات، التكوين المستقبلي للسكان، العدد حسب الجنس والعمر. إن النظر في التغيرات المحتملة في عدد السكان وبنيتهم العمرية يلعب دورًا مهمًا في تحديد الخطوات التي يجب اتخاذها للتنمية المستقبلية. فالتخطيط الرفيع المستوى يمكن أن يبني بشكل فعال بنية تحتية وطنية سليمة ويحسن نوعية الخدمات المقدمة للسكان. بدون التنبؤات السكانية، في مجال تعزيز نظام الحماية الاجتماعية للسكان، والحد من الفقر، ورفع مستوى معيشة السكان، وما إلى ذلك.

لا يمكن وضع استراتيجيات في هذه الاتجاهات، والتخطيط على المدى الطويل التنمية الاجتماعية والاقتصادية المعقدة[2]. كما تُستخدم الإسقاطات الديموغرافية (السكانية) على نطاق واسع في التخطيط الاستراتيجي في القطاع الخاص، من قبل الأكاديميين وغيرهم من الباحثين، وخاصة في مجالات التعليم والصحة والعلوم الاجتماعية [3].

يتم إعداد التنبؤات السكانية الوطنية والدولية حسب حجم السكان والعمر والجنس من قبل العديد من المنظمات ذات السمعة الطيبة في العالم، حيث ان الأمم المتحدة هي المنظمة الرئيسية التي تقوم بإعداد التقديرات والتنبؤات السكانية المحدثة بانتظام[11]. وقدمت تقديرات سكانية لـ 237 دولة وإسقاطات للأعداد حسب العمر والجنس حتى عام 2100، بناءً على تحليل الاتجاهات الديموغرافية التاريخية بدءًا من عام 1950[5]. وقد نُشرت هذه التوقعات منذ عام 1953 في منشور التوقعات السكانية العالمية. تُستخدم التوقعات كبيانات مدخلة في نمذجة القضايا العالمية مثل الأمن الغذائي وتغير المناخ. تستخدم العديد من البلدان أيضًا بيانات التوقعات هذه للتخطيط الوطني [2، 3].

وفي بعض البلدان، تقوم الحكومات الوطنية بإعداد هذه التوقعات. تتضمن هذه التوقعات عادةً تنبؤات للسنوات الأربعين إلى الخمسين القادمة أو نحو ذلك. على سبيل المثال، بالنسبة للقضايا الوطنية الكبرى المتعلقة بالبنية التحتية وتخطيط القوى العاملة، يتوقع مكتب الإحصاء الأمريكي أن عدد سكان الولايات المتحدة سوف ينمو على مدى الأربعين سنة القادمة [3]. تتنبأ الحكومة اليابانية بمستقبل سكانها خلال الخمسين سنة القادمة [4]. تضع إدارة الضمان الاجتماعي الأمريكية ميزانيتها للسنوات الـ 75 القادمة، والتي تتضمن توقعات الوفيات وغيرها من التوقعات السكانية لتلك الفترة المتوقعة [2، 5].

إن التوافر المتزايد لكميات كبيرة من البيانات التاريخية والحاجة إلى التنبؤ الدقيق بالسلوك المستقبلي في عدد من المجالات العلمية والتطبيقية قد استلزم استخدام أساليب موثوقة وفعالة[2]. منذ ستينيات القرن العشرين، لعبت الأساليب الإحصائية الخطية مثل نماذج الانحدار الذاتي للمتوسط المتحرك المتكامل (ARIMA) دورًا مهمًا في مجال التنبؤ[12]. تشير بعض الدراسات إلى أن هذه الأساليب توفر قيمًا تنبؤية أقل بشكل منهجي من الطرق الإحصائية البسيطة وتكون صالحة مع أحجام عينات منخفضة للغاية[6]. في الآونة الأخيرة، اكتسبت نماذج التعلم الآلي الاهتمام وبرزت كمنافس جدي للنماذج الإحصائية الكلاسيكية في مجتمع التنبؤ. على الرغم من أن نموذج ARIMA كان أداؤه أفضل في العديد من الحالات، إلا أن الباحثين خلصوا إلى أن نماذج التعلم الآلي أكثر قدرة على التنبؤ بمجموعات البيانات الأكبر[5]. ويثبت الاستخدام الواسع النطاق لخوارزميات التعلم الآلي في السنوات الأخيرة من الأبحاث حول التنبؤ بالسلاسل الزمنية، وكذلك السلاسل الزمنية الديموغرافية، ذلك مرة أخرى[1].

في السنوات الأخيرة، ظهرت نماذج تعلم الآلة كمنافس جدي للنماذج الإحصائية الكلاسيكية في مجتمع التنبؤ[3]. يستخدم نموذج التعلم الآلي تقنيات تتعلم وتدرب باستخدام البيانات الموجودة لإجراء اتصالات بين بيانات الإدخال والإخراج[2]. يمكن لنموذج تعلم الآلة الذي تم تدريبه بنجاح تحليل البيانات وإجراء تنبؤات صحيحة حتى في حالة عدم توفر البيانات التاريخية لمنطقة الدراسة أو فقدان بعض ميزات الإدخال[10].

ومع ذلك، على الرغم من استخدام بعض خوارزميات التعلم الآلي في دراسات أخرى، إلا أنه لم يتم اختبارها في التنبؤ السكاني في العراق. تم في هذا البحث استخدام تعليم الآلة والمتمثلة بالشبكات العصبية وكذلك أنموذج السلاسل الزمنية ARIMA للتنبؤ السكاني في العراق. اذ ان الهدف من هذا البحث هو بناء أفضل نموذج للتعلم الآلي لبيانات السلاسل الزمنية للسكان وإجراء تنبؤ صحيح بحجم سكان العراق للسنوات الخمس القادمة.

2. الاسقاطات السكانية

نتناول الآن بعض المفاهيم الأساسية للإسقاطات السكانية، تم تحديد طريقة المكونات الأترابية للإسقاط السكاني (CCMPP) لأول مرة بواسطة كانان (1895)، وتم تطويرها بمزيد من التفصيل بواسطة ويلتون (1928،1936)[3]. وأصبحت الطريقة القياسية التي يستخدمها مكتب الإحصاء الأمريكي بدءًا من الأربعينيات، ثم انتشرت بعد ذلك في جميع أنحاء العالم[4]. ويتم استخدامه الآن من قبل معظم الوكالات الوطنية التي تنتج التنبؤات السكانية، إنها طريقة حتمية (أي غير احتمالية) [7]. وفي الوقت نفسه، تم إضفاء الطابع الرسمي على الطريقة رياضيًا بواسطة ليزلي (1945)، باستخدام ما أصبح يعرف باسم مصفوفة ليزلي كمفهوم أساسي لها[1]. للحصول المكونات الثلاثة للتغير السكاني هي الولادات والوفيات والهجرة، وتحدث هذه في وقت مستمر ولكن يتم تجميعها عادةً في فترات زمنية[11]. ونتيجة لذلك، تشير معظم البيانات الديموغرافية (أعداد المواليد والوفيات وأحداث الهجرة) إلى فترات زمنية منفصلة مثل السنوات أو فترات الخمس سنوات[10]. تعتبر الرياضيات الخاصة بعمليات الزمن المستمر الأساسية بسيطة نسبيًا، وإن طرق تحليل البيانات الزمنية المنفصلة المتاحة للعلماء الديموغرافيين تنطوي حتما على تقديرات تقريبية يمكن أن تكون معقدة إلى حد ما[4]. وان من اهم المؤشرات السكانية هي; معدل الوفيات، معدل الخصوبة و الهجرة[6].

3. تعليم الآلة

قبل بدء الشرح لأساليب التعلم الآلي، لابد ان يكون لدينا فهم جيد لماهية التعلم الآلي[9]. ويمكن تعريف تعلم الآلة على انه فرع من الذكاء الاصطناعي، الذي يستخدم مجموعة متنوعة من الأدوات الإحصائية والاحتمالية والتحسينية "للتعلم"[5]. من الأمثلة السابقة ثم استخدام هذا التدريب المسبق لتصنيف البيانات الجديدة أو تحديد أنماط جديدة أو التنبؤ بالاتجاهات الجديدة [3]. يُستخدم التعلم الآلي، مثل الإحصائيات، لتحليل البيانات وتفسيرها[9]. على عكس الإحصائيات، يمكن أن تستخدم أساليب التعلم الآلي المنطق البولياني (AND، OR، NOT)، والشرطية المطلقة (IF، THEN، ELSE)، والاحتمالات الشرطية (احتمال X مع Y) واستراتيجيات التحسين غير التقليدية لنمذجة البيانات أو تصنيف الأنماط . تشبه هذه الأساليب الأخيرة في الواقع الأساليب التي يستخدمها البشر عادةً للتعلم والتصنيف[11]. لا يزال التعلم الآلي يعتمد بشكل كبير على الإحصائيات والاحتمالات، لكنه أقوى بشكل أساسي لأنه يسمح باتخاذ استنتاجات أو قرارات لا يمكن اتخاذها باستخدام المنهجيات الإحصائية التقليدية[6]. على سبيل المثال، تعتمد العديد من الأساليب الإحصائية على تحليل الانحدار أو الارتباط متعدد المتغيرات على الرغم من كونها قوية جدًا بشكل عام، إلا أن هذه الأساليب تفترض أن المتغيرات مستقلة وأن البيانات يمكن نمذجتها باستخدام مجموعات خطية من هذه المتغيرات[12]. عندما تكون العلاقات غير خطية وتكون المتغيرات مترابطة (أو تعتمد بشكل مشروط)، فإن الإحصائيات التقليدية عادة ما تتعثر. في هذه الحالات يميل التعلم الآلي الى النماذج غير خطية[4].

ان تعليم الالة يستند على خطوات اساسية في عمله والتي هي[2,8,10]:

- ✓ **اختيار النموذج:** هناك أنواع مختلفة من نماذج التعلم الآلي، كل منها مناسب لأنواع مختلفة من المهام.
- ✓ **تدريب النموذج:** خلال مرحلة التدريب، يتم تغذية نموذج التعلم الآلي المحدد بالبيانات المعدة. يقوم النموذج بضبط معلماته الداخلية بشكل متكرر لتقليل الفرق بين تنبؤاته والقيم الفعلية في بيانات التدريب. تتم هذه العملية عادةً باستخدام خوارزميات التحسين.

- ✓ **تقييم النموذج:** بمجرد تدريب النموذج، يجب تقييمه لتقييم أدائه على البيانات غير المرئية. ويتم ذلك عادةً عن طريق تقسيم البيانات المتاحة إلى مجموعات تدريب واختبار. تعتمد مقاييس التقييم الشائعة على نوع المشكلة التي يتم حلها، مثل الدقة أو الدقة أو الاستدعاء أو درجة F1 للتصنيف أو متوسط الخطأ التربيعي أو R^2 للانحدار.
- ✓ **ضبط النموذج:** اعتمادًا على أداء النموذج، قد يحتاج إلى ضبط دقيق عن طريق ضبط المعلمات أو حتى عن طريق تحديد خوارزمية مختلفة. تسمى هذه العملية ضبط المعلمة ويتم إجراؤها غالبًا باستخدام تقنيات تحدد.

4. الشبكات العصبية

الشبكات العصبية الاصطناعية انتشرت في السنوات الأخيرة في العديد من المجالات؛ يمكن تعريف ANN بأنها شبكات معالجات بسيطة تتكون من وحدات وعقد وخلايا عصبية. وترتبط هذه المعالجات بوصلات اتصال أحادية الاتجاه تحمل بيانات رقمية[1]. إنه يعمل مثل الدماغ من ناحيتين؛ يتم استخدام اتصال نقاط قوة Interneuron المعروفة باسم الأوزان المشبكية لتخزين المعرفة وتم الحصول على المعرفة بواسطة الشبكة من خلال عملية التعلم [3].

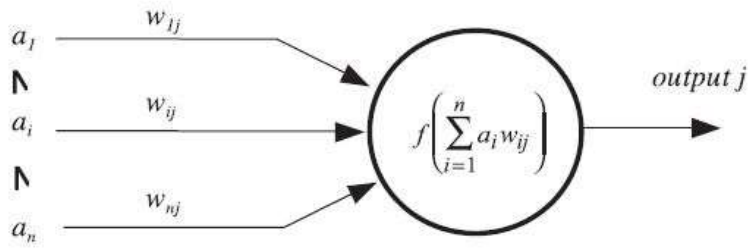
تم تصميم ANN على شكل صندوق أسود يحتوي على مدخلات وينتج مخرجات؛ وتعتمد وظيفة الصندوق الأسود على بنية الشبكات العصبية وجميع الخلايا العصبية في هذا الهيكل للنموذج[4]. جاء الاهتمام بالشبكة العصبية الاصطناعية (ANN) من قدرة دماغ الإنسان على تنظيم الخلايا العصبية لتنفيذ عمليات حسابية معينة، مثل التعرف على الأسلوب، والتمييز، وبسرعة تفوق سرعة أسرع كمبيوتر رقمي في هذه الأيام[1]. تستمد الشبكة ANN قوتها الحسابية من خلال بنيتها الموزعة المتوازية الضخمة وقدرتها على التعلم والتعميم[2]. ويشير التعميم إلى قدرة الشبكة العصبية الاصطناعية على تقديم استجابات كافية للمدخلات، وهو ما اختفت أثناء تدريبها ويمكننا أن نقول عملية "التعلم"، تحتوي نماذج ANN على خصائص مفيدة[3]. يتم سرد هذه الخصائص على النحو التالي: اللاخطية، والتكيف، والقدرة على تنفيذ VLSI، والقياس العصبي البيولوجي [2]. بشكل عام، تتكون الشبكة العصبية الاصطناعية من ثلاثة مكونات: طوبولوجيا الشبكة، والخلايا العصبية، وخوارزمية التعلم[11]. الخلايا العصبية هي المكونات التي يتم فيها الحساب في أنظمة ANN. يوضح الشكل1 أن الخلايا العصبية تتلقى إشارات من مدخلات أخرى من خلال الارتباطات. يمكن صياغة هذه الإشارات كمتجه للمدخلات.

$$A= (a_1, L, a_i, L, a_n) \quad (1)$$

الارتباطات مع كل زوج من الخلايا العصبية المتصلة هي قيمة قابلة للتعديل تسمى الوزن. ترتبط هذه الأوزان بالخلية العصبية j ويمكن صياغتها كمتجه للشكل.

$$W_j = (w_{1j}, L, w_{ij}, L, w_{nj}) \quad (2)$$

حيث يمثل Wij قوة الاتصال من الخلية العصبية a_i إلى الخلية العصبية a_j



شكل (1): يمثل العقدة في الشبكة العصبية

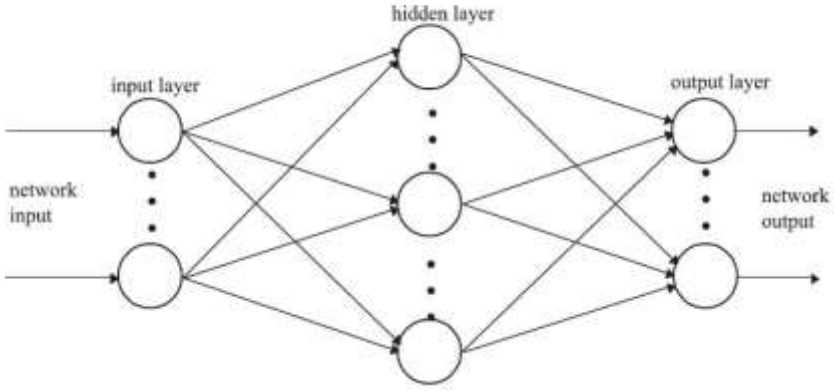
للخلية العصبية حد محلي، وعندما تتجاوز هذه الحدود عتبة، يمكنها نشر جهة اتصال من خلال اتصالاتها مع الخلايا العصبية الأخرى. يمكن أن تكون الخلايا العصبية نشطة في نفس الوقت، وبالتالي يمكن اعتبار الشبكات العصبية أنظمة حسابية متوازية [2]. يمكن تعريف قيمة الإخراج للخلية العصبية j على النحو التالي:

$$\text{output}_j = f\left(\sum_{i=1}^{n} aiwij\right) \quad (3)$$

هناك العديد من الوظائف التي يمكن استخدامها. جزء من هذه الوظائف هي غير الخطية، الخطية، الغوسية، والسينية. تعد الدالة السيني خيارًا شائعًا ويمكن كتابتها بواسطة:

$$f(x) = \frac{1}{1+e^{-x}} \quad (4)$$

يتم ربط الخلايا العصبية ANN معًا ويتم الحصول على سلوك النظام الإجمالي من خلال بنية وأوزان هذه الوصلات. يتم ترتيب الخلايا العصبية في طبقات. يمكن أن تكون هذه الطبقات عبارة عن طبقات إدخال أو طبقات مخفية أو طبقات إخراج. توفر الطبقات المخفية للشبكات القدرة على إجراء تعيينات غير خطية[w]. يمكن أن تنتشر الإشارات بين الطبقات بطريقتين؛ تغذية إلى الأمام وإشارات ردود الفعل. يظهر الهيكل العام لـ ANN في الشكل 2.



شكل (2): يظهر الإدخال والطبقات المخفية وطبقة الإخراج للشبكة العصبية

5. ARIMA

تعتبر السلاسل الزمنية من الاساليب الاحصائية المهمة بسبب طبيعتها الديناميكية في نمذجة البيانات, فهي تجمع البيانات بناءً على المشاهدات السابقة[12]. التنبؤ الناجح بالسلاسل الزمنية هو عملية التنبؤ بالمستقبل من خلال تحليل قيم المشاهدات السابقة للحالة قيد الدراسة. ان أحد أشهر نماذج السلاسل الزمنية العشوائية هو ARIMA [6]. تم بناء ARIMA من خلال ثلاثة نماذج فرعية: الانحدار الذاتي (AR)، والمتوسط المتحرك (MA)، وARMA. كما أنه مفيد جدًا للتنبؤ المستند إلى السلاسل الزمنية الموسمية[12]. يتم تقدير معلمات السلاسل الزمنية باستخدام الترتيب المناسب لـ ARIMA (p، d، q) في المرحلة الثانية من النموذج. تم إعداد ACF (دالة الارتباط الذاتي) وPACF (دالة الارتباط الذاتي الجزئي) بناءً على معلمات p وd وq. المعادلات الأساسية هي كما يلي لـ ARIMA

$$\text{AR}(p): xt = a_0 + a_1x_{t-1} + a_2x_{t-2} + \cdots + a_px_{t-p} + e_t \quad (5)$$

$$\text{MA}(q): x_t - a_0 = e_t + b_1e_{t-1} + b_2e_{t-2} + \cdots + b_qe_{t-q}, \quad (6)$$

يتم التنبؤ ببيانات السلاسل الزمنية من نموذج ARIMA. حيث تُستخدم دالة التنبؤ للتنبؤ بدالة تركيب النماذج المختلفة. أثناء إجراء ARIMA، يتم اختيار عدد من النماذج بناءً على قيم التأخر (Lags) على AR وMA. يتم تقييم هذه النماذج ويتم تحديد النموذج النهائي بناءً على معيار MAPE اذ يعتبر أدنى قيمة له أقرب إلى البيانات الحقيقية و كذلك على معيار R^2 حيث كلما اقترب قيمته من الواحد الصحيح دل على جودة النموذج, ويتم استخدامه ايضاً لقياس جودة البيانات الإحصائية [8]. حيث ان الصيغ الرياضية لهذه المعيارين كالاتي:

$$MAPE = (1/N) \sum (|y_i - \hat{y_i}| / y_i) \quad (7)$$

$$R^2 = 1 - (\sum (\hat{y_i} - y_i)^2 / \sum (\bar{y} - y_i)^2) \quad (8)$$

حيث ان

y_i : تمثل القيم الحقيقة للظاهرة المدروسة

$\hat{y}$: تمثل القيم التقديرية للظاهرة المدروسة

$\bar{y}$: يمثل الوسط الحسابي للقيم الحقيقة للظاهرة المدروسة

6. جمع البيانات

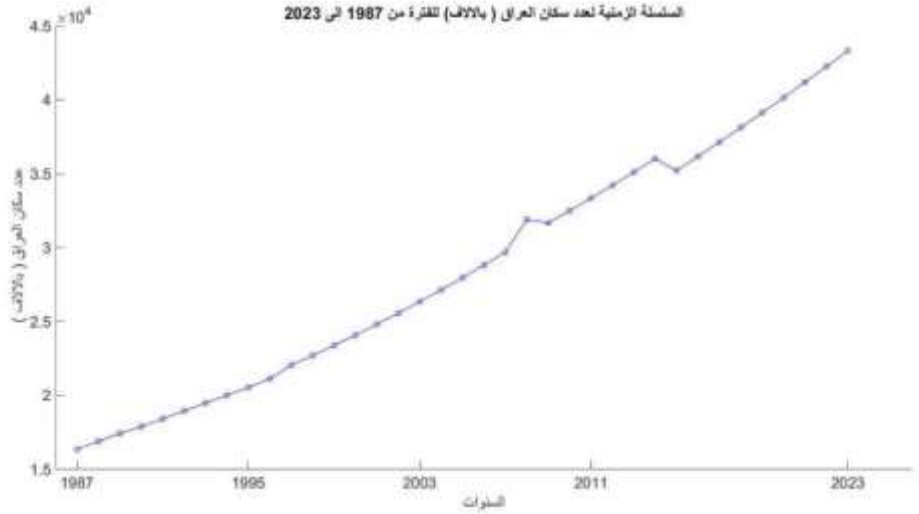
تم في هذا البحث جمع اعداد سكان العراق من عام 1987 والذي يعتبر التعداد العام لسكان العراق قبل الاخير, اذ ان اخر تعداد عام لسكان تم اجراءه في العراق كان في سنة 1997. حيث تم الاعتماد على نشرات وزارة التخطيط/الجهاز المركزي للإحصاء، حيث كانت البيانات المسجلة من عام 1987 الى عام 2023.

جدول (1): يمثل بيانات سكان العراق(بالالاف)

التسلسل	السنة	العدد(بالالاف)	التسلسل	السنة	العدد(بالالاف)
1	1987*	16355	20	2006	28810
2	1988	16892	21	2007	29682
3	1989	17428	22	2008	31895
4	1990	17890	23	2009**	31664
5	1991	18419	24	2010***	32490

33338	2011***	25	18949	1992	6
34208	2012***	26	19478	1993	7
35096	2013***	27	20007	1994	8
36005	2014***	28	20536	1995	9
35213	2015***	29	21124	1996	10
36169	2016***	30	22046	1997*	11
37140	2017***	31	22702	1998	12
38124	2018***	32	23382	1999	13
39128	2019***	33	24086	2000	14
40150	2020***	34	24813	2001	15
41190	2021***	35	25565	2002	16
42248	2022***	36	26340	2003	17
43324	2023***	37	27139	2004	18
			27963	2005	19

*حسب نتائج التعداد العام للسكان , **حسب نتائج الترقيم والحصر لعام , ***محتسبة حسب فرضيات سكانية جديدة (2010-2023).



شكل (3): يبين عدد سكان العراق(بالالاف) من1987-2023

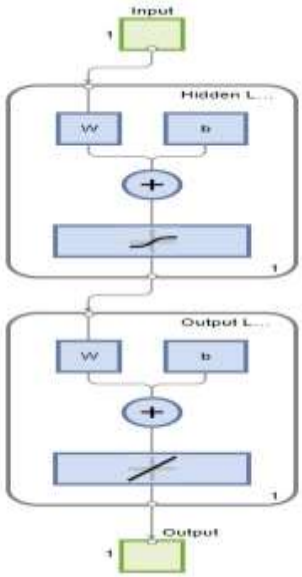
7. النتائج ومناقشتها

في هذا القسم، تم دراسة القدرة التنبؤية لخوارزمية تعلم الآلة والمتمثلة (ANN) المستخدمة في الدراسة وتمت مقارنة النتائج. باستخدام التحديد العشوائي، يتم تقسيم مجموعة بيانات التدريب الأصلية إلى مجموعات بيانات تدريب واختبار لنماذج التعلم الآلي. وهذا ضروري لتدريب النماذج ومقارنتها واختبارها. تم استخدام 70% من البيانات لتدريب النماذج و30% لبيانات الاختبار. تم إجراء تحليل البيانات باستخدام لغة البرمجة Matlab2022A حيث كانت النتائج التدريب والاختبار كما موضح في الجدول2.

جدول (2): يمثل نتائج الشبكة العصبية

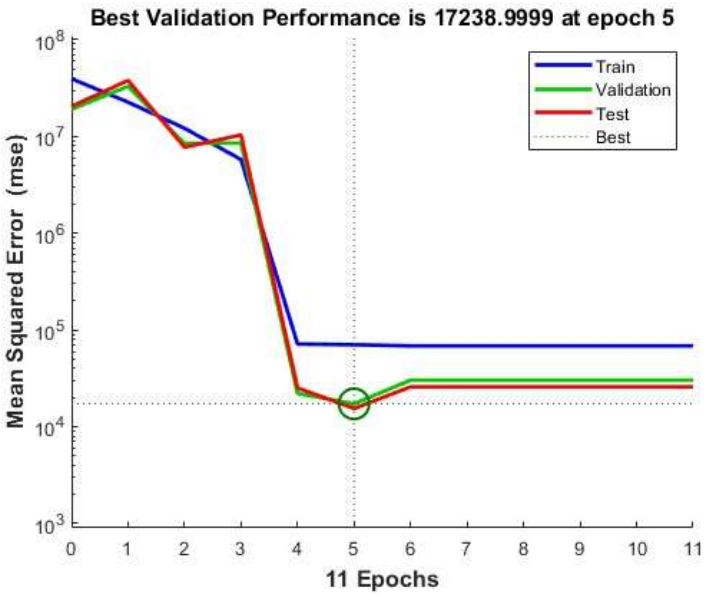
عدد المشاهدات التدريب	عدد المشاهدات الاختبار	عدد الطبقات	عدد العقد المستخدمة	MAPE (%)	R^2 (%)
27	10	1	100	0.0673414	0.9644
			150	0.0512402	0.9982
			200	0.0601213	0.9752

اذ كان عدد مشاهدات التدريب 27 مشاهدة وعدد مشاهدات الاختبار 10, اما عدد الطبقات التي استخدمت هو 1 في حين تم استخدام عدد عقد (100, 150 و 200). ومن خلال نتائج الشبكة العصبية نلاحظ ان اقل قيمة لمعيار MAPE كانت عند عدد عقد 150 حيث بلغت (0.512402) في حين كانت اعلى قيمة لR^2 عند نفس عدد العقد اذ بلغت (0.9982). كما موضح بالاشكال التالية:



شكل (4): يمثل معمارية الشبكة العصبية

ومن خلال الشكل اعلاه نلاحظ إن متغيرات الإدخال مرتبطة بعقد الطبقة المخفية والتي عددها (150) عقدة وكلاهما مرتبطتان بعقدة الإخراج وان قوة العلاقة ما بين العقد سواء كانت للإدخال والإخراج يتحدد من خلال قيم الاوزان (W_{ij}) والتي تعطى بشكل ابتدائي لكي يتم تدريب الشبكة عليها من خلال عملية الانبعاث الخلفي للبيانات.



شكل (5): يمثل التحقق الشبكة العصبية

من خلال الشكل اعلاه نلاحظ ان افضل معدل خطا توصلت الي هو عند النقطة 10^4 من خلال عملية تدريب اوزان الشبكة العصبية والتي تحققت عند (1723.99) عند الدورة الخامسة وان العدد الكلي للدورات حتى التوقف هو 11.

وان القيم التنبؤية التي لعدد سكان العراق لخمسة سنوات قادمة كانت:

جدول (3): يمثل القيم التنبؤية لسكان العراق (بالألاف) لخمسة سنوات قادمة من خلال الشبكات العصبية

التسلسل	السنة	العدد(بالالاف)
1	2024	42775.6865
2	2025	43298.7062
3	2026	43791.6454
4	2027	44255.1447
5	2028	44690.0010

حيث نلاحظ من خلال نتائج التنبؤ بالشبكة العصبية ان عدد سكان العراق سيزداد ليصل في عام 2028 الى 44690.0010 (بالالف) وكما موضح بالشكل6

شكل (6): يمثل التنبؤ بعدد سكان العراق للفترة (2024-2028) من خلال الشبكة العصبية

ومن ايجاد افضل نموذج ARIMA يتلاءم مع البيانات تم الاعتماد على قيم المعايير الاحصائية (MAPE,R^2) وعلى مصفوفة الارتباط الذاتي ومصفوفة الارتباط الذاتي الجزئي , وليكن فحصه دقيق فسنقوم بعمل جدول يمثل المعايير الاحصائية للنموذج الافضل

جدول (4): يمثل نتائج نماذج ARIMA

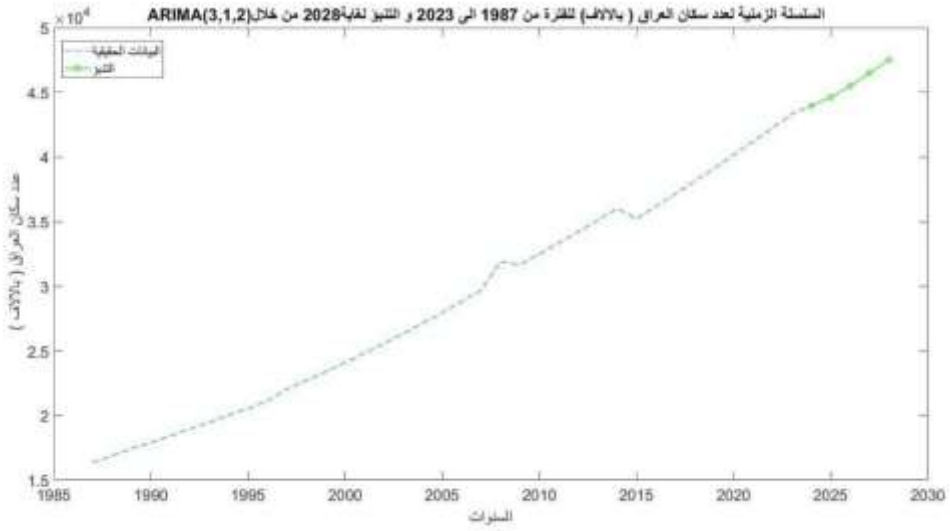
Models	MAPE%	R^2%
ARIMA(1,1,0)	0.1271	0.9943
ARIMA(2,1,0)	0.1183	0.9953
ARIMA(3,1,0)	0.1109	0.9958
ARIMA(1,1,1)	0.0884	0.9968
ARIMA(2,1,1)	0.0838	0.9973
ARIMA(3,1,1)	0.1125	0.9953
ARIMA(3,1,2)	0.0704	0.9975
ARIMA(1,2,1)	0.0807	0.9970

من خلال جدول 4 وبعد حساب قيمة المعايير نلاحظ ان النموذج ARIMA(1,2,1) هو الاكثر ملائمة للبيانات حيث بلغ قيمة MAPE هو 0.0704 و قيمة R^2 0.9975.

جدول (5): يمثل القيم التنبؤية لسكان العراق(بالالاف) لخمسة سنوات قادمة من خلال ARIMA

التسلسل	السنة	العدد (بالألاف)
1	2024	43960.2781
2	2025	44618.2294
3	2026	45493.0341
4	2027	46497.3235
5	2028	47516.3113

حيث نلاحظ من خلال نتائج التنبؤ بالشبكة العصبية ان عدد سكان العراق سيزداد ليصل في عام 2028 الى 47516.3113 (بالألف) وكما موضح بالشكل 7



شكل (7): يمثل التنبؤ بعدد سكان العراق للفترة (2024-2028) من خلال نموذج ARIMA(3,1,2)

جدول (6): يمثل مقارنة القيم التنبؤية لسكان العراق(بالألاف) لخمسة سنوات قادمة

التسلسل	السنة	الشبكات العصبية العدد (بالألاف)	ARIMA(3,2,1) العدد (بالألاف)
1	2024	42775.6865	43960.2781
2	2025	43298.7062	44618.2294
3	2026	43791.6454	45493.0341
4	2027	44255.1447	46497.3235
5	2028	44690.0010	47516.3113

8. الاستنتاجات

من خلال ما تقدم من نتائج من خلال استخدام اسلوبين في التنبؤ بعدد سكان العراق لخمسة سنوات قادمة, نلاحظ تفوق تعليم الالة المتمثل بالشبكات العصبية ذات الانبعاث الخلفي, اذ كانت تمتلك اقل MAPE وعلى اعلى R^2, اذ ان نتائج الشبكة العصبية تنبأت بارتفاع عدد سكان العراق الى ما يقارب 44690.0010 (بالألاف) في عام 2028.

المصادر

[1] Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H. and Saidur, R. (2014) 'A review on applications of ANN and SVM for building electrical energy consumption forecasting', Renewable and Sustainable Energy Reviews, Vol. 33, No. 1, pp.102–109.

[2] Alkema, L., Raftery, A.E., Gerland P., Clark, S.J. and Pelletier, F. (2012) 'Estimating trends in the total fertility rate with uncertainty using imperfect data: examples from west Africa', Demographic Research, Vol. 26, No. 15.

[3] B. Nazlı, Y. Gültepe, and H. Altural, "Classification of Coronary Artery Disease Using Different Machine Learning Algorithms," International Journal of Education and Management Engineering (IJEME), Vol.10, No.4, pp.1-7, 2020. DOI: 10.5815/ijeme.2020.04.01

[4] Bandyopadhyay, G. and Chattopadhyay, S. (2006) An Artificial Neural Net Approach to Forecast the Population of India, arXiv preprint nlin/0607058.

[5] Bélisle, E., Huang, Z., Le Digabel, S. and Gheribi, A.E. (2015) 'Evaluation of machine learning interpolation techniques for prediction of physical properties', Computational Materials Science, Vol. 98, pp.170–177.

[6] Boon, M.E. and Kok, L.P. (1995) 'Classification of cells in cervical smears', Applications of Neural Networks, Springer, Boston, MA, pp.113–131.

[7] C.Y. Wang and S.J. Lee, "Regional Population Forecast and Analysis Based on Machine Learning Strategy," Entropy, vol. 23, no. 656, pp. 1-12, 2021. Available at: https://doi.org/10.3390/e23060656

[8] D. Chicco, M.J. Warrens, G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," Peer J Computer Science 7:e623, 2021,https://doi.org/10.7717/peerj-cs.623

[9] Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange." Expert Systems with Applications 38(5): 5311-5319.

[10] Minal Patel, Sanjay Chaudhary, Sanjay Garg, "Machine Learning Based Statistical Prediction Model for Improving Performance of Live Virtual Machine Migration", Journal of Engineering, vol. 2016, Article ID 3061674, 9 pages, 2016. https://doi.org/10.1155/2016/3061674Vol. 3, No. 8, p.311.

[11] S. B. Rajakumari, P. Padmanabhan, S. Christy, and M. Nandhini, "Prediction of population growth using machine learning techniques," European Journal of Molecular & Clinical Medicine, Vol.7, No. 5, pp. 1871-1879, 2020.

[12] V. S. Fatih, T.T. Ahmet, and C. Ferhan, "Machine Learning algorithm to forecast the population: Turkey Example," in Proceedings of International Engineering and Technology Management Summit 2019 – ETMS. Available at: www.researchgate.net/publication/33714439

Journal of AL-Rafidain University College for Sciences (2024); Issue 56; pp. 88-97

AL- Rafidain University College

PISSN: (1681-6870); EISSN: (2790-2293)

Journal of AL-Rafidain University College for Sciences

Available online at: https://www.jrucs.iq

JRUCS

Journal of AL-Rafidain University College for Sciences

Using Machine Learning in Statistical Models for Population Forecasting in Iraq

Omar A. Abd Alwahab	Maryam M. Salman
omaradil.d87@gmail.com	maryam90salman@gmail.com
Statistic Department, College of Administration and Economics, University Of Diyala, Diyala, Iraq	Ibn Sina University of Medical and Pharmaceutical Sciences, Baghdad, Iraq

Article Information

Article History:
Received: February, 18, 2024
Accepted: April, 12, 2024
Available Online: December, 31, 2024

Keywords:
Population projections, machine learning, neural networks, ARIMA, prediction.

Correspondence:
Omar A. Abd alwahab
omaradil.d87@gmail.com

Abstract

Statistical studies are the primary focus of the independent planning method in any country, as the state and the private sector use population projections in planning. Another essential element is the provision of comprehensive services, which requires knowing how many people there will be in the future, and what they are divided by age, gender and characteristics, such as ethnicity. Geography, etc. Population forecasts are also widely used in the private sector for strategic planning, as population projections are considered one of the most important indicators used by its maker and population analyzes over time. In this research, the machine Learning by using neural networks, in addition to using a classical method, which is the ARIMA time series model, to predict population projections. The internal data was scanned in both methods, up to 70% training and 30% testing, and a five-year forecast for the Iraqi population. The research found that machine learning has an advantage in estimation because it has the lowest MAPE and higher R2. The population forecast of Iraq in 2028 is 47,516.3113 (per thousand).

DOI: https://doi.org/10.55562/jrucs.v56i1.8