

## طريقة مقترحة للتنبؤ بالإيرادات الضريبية باستخدام خوارزميات الذكاء الاصطناعي

## A proposed method for forecasting Tax Revenue Using Artificial Intelligence Algorithms

أنور حسان محمود

قسم الإصلاح المصرفي/ الدائرة الاقتصادية

[Anwer.h.mahmood@gmail.com](mailto:Anwer.h.mahmood@gmail.com)

## المستخلص (Abstract):

تعتبر الضرائب أحد المصادر الرئيسية لتمويل ميزانية الحكومة. لذلك، فإن وجود صورة واضحة عن المبلغ الذي يمكن تحقيقه من الضرائب ليس ضروريًا فقط للتخصيص الأمثل للموارد، ولكنه يساعد أيضًا الحكومة على تطوير برامج دقيقة لتحصيل الضرائب. في هذه الدراسة يتم توظيف خوارزمية التحسين الجسيمات (PSO) Particular Swarm Optimization والذاكرة الطويلة قصيرة المدى (LSTM) Long Short Term Memory لغرض التنبؤ بالتدفقات النقدية الضريبية، باستخدام البيانات للفترة 2003-2022، اظهرت معايير الدقة للانموذج ويمكن الاعتماد عليه من قبل متخذي القرار، وبينت النتائج أن الانموذج المقترح (LSTMPSO) يتفوق على إنموذج (LSTM) في التنبؤ بإيرادات الضرائب وذلك بمتوسط خطأ نسبي مطلق MAPE (0.159828).

الكلمات المفتاحية: الإيرادات الضريبية، خوارزميات الذكاء الاصطناعي .

## 1-1 المقدمة (INTRODUCTION):

إن تمويل الموازنة الحكومية من خلال تحصيل الإيرادات الضريبية هو أحد الأدوات اعداد الموازنة العامة . إن الحصول على صورة واضحة حول المبلغ الذي يمكن تحقيقه من الضرائب ليس ضروريًا فقط للتخصيص الأمثل للموارد النادرة لتحصيل الضرائب، ولكنه يساعد أيضًا الحكومة على تطوير برامج دقيقة لتحصيل الضرائب. نظرا للخصائص المتأصلة للمتغيرات الاقتصادية وأيضاً لتأثير العوامل الاجتماعية والسياسية والنفسية والبيئية على المتغيرات الاقتصادية، يتم وضع تقديرات الإيرادات الضريبية الموازنة العامة بالإعتماد على المبالغ التي يتم تقديمها من الوزارات والجهات غيرالمرتبطة بوزارة وغالباً ما يتم تقديم هذه المبالغ إستناداً على بيانات السنة السابقة مع إضافة مبالغ معينة يتم إضافتها بشكل غير دقيق ومبالغ فيه وهذا ما يتم ملاحظته للمدة الزمنية من 2003 ولغاية 2023 لهذا السبب أغلب الموازنات إما تنتهي بحالة عجز أو فائض وذلك بسبب عدم التنفيذ والتخطيط والتقدير بصورة صحيحة [1]. يعتبر التنبؤ بالتدفقات النقدية الضريبية هو عنصر أساسي ومهم لعملية إعداد الموازنة العامة في حال وضع تقدير مُبالغ فيه للإيرادات والحصول على تدفقات نقدية أقل من التقديرات المتوقعة فإن ذلك يؤدي الى فشل البرامج الاقتصادية والاجتماعية وخصخصة الأصول الثابتة وزيادة الضرائب وذلك لغرض تسوية المغالات [2]. هنالك اسلوبان يتم استخدامهما في وضع تقديرات الإيرادات الضريبية هما الأسلوب النوعي والكمي [3]، غالباً ما يكون لأسلوب النوعي مُعرض الى التحيز والتحفز، بسبب اعتمادها على التحكيم الشخصي وتكون شائعة الاستخدام في حال البيانات التاريخية غير موجودة، بينما يعتبر الاسلوب الكمي اسلوب علمي يتم الاعتماد عليها بإستشراق المستقبل بإستخدام و تحليل المتسلسلات الزمنية والتنبؤ. في هذه الدراسة تم إستخدام خوارزمية التحسين الجسيمات (PSO) والذاكرة طويلة قصيرة المدى (LSTM) للتنبؤ بالإيرادات الضريبية.

## 1-2 مشكلة الدراسة (Problem statement):

تَكمن مشكلة الدراسة في أنَّ التدفقات النقدية الضريبية تعاني من إختلاف بين الإيرادات الضريبية التخطيطية والإيرادات الضريبية الفعلية المتحصل عليها، بسبب عدم استخدام الأسس العلمية الصحيحة والدقيقة عند وضع تقديرات الإيرادات الضريبية مما يسبب فجوة حقيقية بين التقديرات التخمينية والتنفيذ الفعلي.

## 1-3 هدف الدراسة (Study Objective):

تَهْدَف هذه الدراسة الى استخدام أساليب الذكاء الاصطناعي ، الذاكرة الطوية قصيرة المدى (LSTM) و خوارزمية التحسين الجسيمات (PSO) وذلك لغرض تحسين إداء الانموذج و لغرض التنبؤ (الإيرادات الضريبية) .

## 1-4 الدراسات السابقة:

هنالك العديد من الدراسات السابقة التي تناولت التنبؤ بالإيرادات والسلاسل الزمنية بشكل عام، على سبيل المثال لا الحصر: أ. في عام 2017 استخدم الباحث (C. Chimilila) إنموذج الإنحدار الذاتي للأوساط المتحركة (ARIMA) والتباين الذاتي المشروط المعمم (GARCH) ، للتنبؤ بإيرادات الضرائب في تنزانيا، بإستخدام بيانات شهرية لمدة 182 شهراً من يناير 2000 إلى فبراير 2015. أظهرت النتائج أن إنموذج (GARCH) يتفوق على إنموذج (ARMA) [3].  
ب. في عام (2019) قدمت الباحثة (K.IBTESAM) دراسة مقارنة بين التنبؤ الديناميكي والتنبؤ الساكن بواسطة تطبيق (ARIMA) للتنبؤ بالفائض والعجز في العراق وأظهرت النتائج تفوق الإنموذج (الديناميكي) على الإنموذج (الساكن) [4].  
ج. وفي نفس العام، قدم (S. Hamidii and A. Hamid) دراسة للتنبؤ بالإيرادات الضريبية في الجمهورية الاسلامية في إيران. بإستخدام إنموذج الشبكة العصبية الاصطناعية ذات التغذية الامامية (FFANN)، أظهرت النتائج أن الإنموذج قدم نتائج عالية الدقة ، في التنبؤ (إجمالي الضريبة، الضريبة المباشرة، الضريبة غير المباشرة، ضريبة الدخل، ضريبة الشركات، ضريبة الأعمال، الضريبة العقارية، ضريبة الرواتب، ضريبة الثروات) ويمكن الاعتماد عليه من قبل متخذي القرار [5].  
د. في عام (2020)، قدم (B. Kozicki et al) دراسة لغرض التنبؤ بالإيرادات في جمهورية بولندا بإستخدام (ARIMA) وإنموذج Holt-Winters للتمهيد الآسي (HW) ، وذلك بإستخدام البيانات من 2011 إلى 2018. وتظهر النتائج أن نماذج (ARIMA) أكثر دقة من (التمهيد الآسي) [6].  
هـ. في (2020)، قدم الباحث (M. C. Technology) دراسة لغرض التنبؤ بالإيرادات السنوية في تنزانيا بالإعتماد على العديد من المتغيرات (الظروف الجوية الزراعية، وسعر الصرف، والناجح المحلي الإجمالي الوطني، وعدد سكان المجلس، وعدد مؤسسات المجلس، والتحصيل السنوي السابق، والشخص المادي ضريبة الدخل) ، وذلك بإستخدام (FFANN) ودعم الانحدار المتجه (SVR). تظهر النتيجة النهائية دقة إنموذج (SVR) التي تم الحصول عليها (94.2%) مقارنة بدقة إنموذج ANN (85%) [2].  
و. وفي نفس السنة ، قدم الباحثون (M. Rhanoui et al) ، دراسة مقارنة بين إنموذجي التنبؤ (ARIMA) مع الذاكرة الطويلة قصيرة المدى (LSTM) للتنبؤ بإيرادات الموازنة العامة في المغرب. حصل إنموذج (LSTM) على دقة أعلى من إنموذج (ARIMA). بسبب قدرة إنموذج LSTM على التعرف على العلاقات الغير خطية بين البيانات [7].

## أولاً: الجانب النظري :

في هذا الجانب سيتم تعريف متغيرات الدراسة بالإضافة الى النماذج الرياضية ونماذج الشبكات العصبية التي سيتم إستخدامها في هذه الدراسة خوارزمية التحسين الجسيمات (PSO) والذاكرة طويلة قصيرة المدى (LSTM).

## 1- الإيرادات العامة (Public Revenue):

تشير إلى مقدار الدخل أو التدفقات النقدية التي تكسبها الحكومة عبر الخزينة العامة، وتشمل الإيرادات السيادية (الضرائب، الرسوم، الكمارك، أخرى). وكذلك الإيرادات النفطية والمعادن وتعتمد عليها الكثير من الدول النامية [2].

## 2- خوارزمية التحسين الجسيمات (PSO) Particular Swarm Optimization:

هي إحدى الخوارزميات التطويرية اقترحت من قبل (James Kennedy & Russell) في عام 1995 [8]. فكرة الخوارزمية مستوحاة من السلوك الاجتماعي لأسراب الطيور أو قطع الأسماك في البحث عن الطعام. وعلى الرغم من عدم معرفة أي طير من الطيور مكان الطعام في البداية إلا أنهم قادرون على إكتساب المعرفة بعد كل تكرار، يشارك كل طائر المعلومات مع باقي أفراد السرب لذا فإن الطائر الذي يبحث عن الطعام في التكرار التالي يعتمد على التجربة الذاتية وتجارب الطيور الأخرى، وبالآلية نفسها تعمل خوارزمية (PSO) على التوصل إلى الحل الأمثل لمشاكل التحسين، تعتمد هذه الخوارزمية على مجموعة من الجسيمات يمثل كل منها حل لمشكلة التحسين، كما تبحث عن القيمة المثلى بعد كل تكرار بينما يتم تقييم الحل الأمثل من طريق دالة اللياقة، كل جسيم في السرب يمتلك موقعا يشار إليه (x) وسرعة يشار إليها (v)، يتم تحديث الموقع حسب (p\_best) الذي يشير إلى أفضل موقع حالي للجسيم الذي تم الحصول عليه من قبل (g\_best) الذي يشير إلى أفضل موقع بين جميع الجسيمات وكما في العلاقتين (1) و(2) الاتينتين [9]:

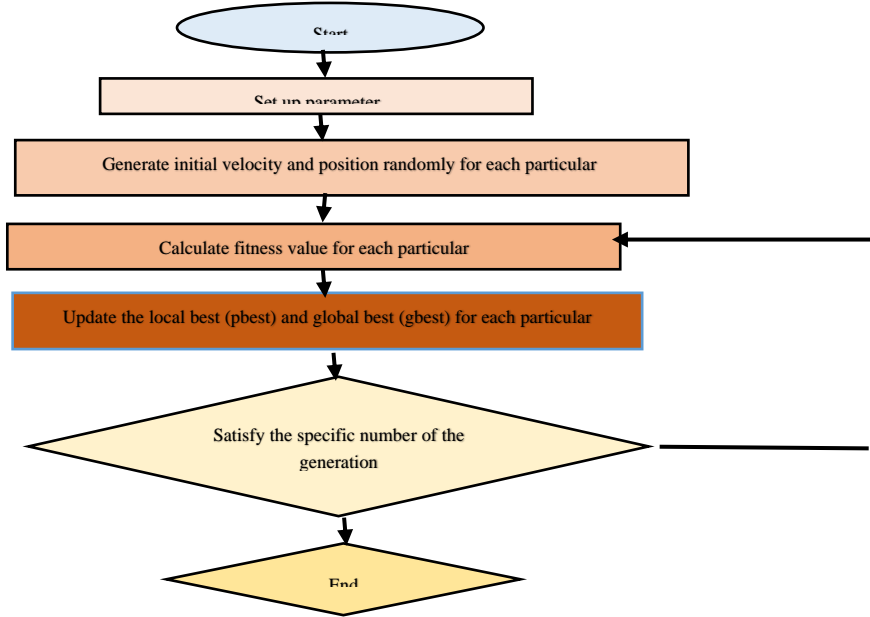
$$v_{t+1} = wv_t + c_1r_1(p\_best - x_t) + c_2r_2(g\_best - x_t) \dots \dots \dots (1)$$

$$x_{t+1} = x_t + v_{t+1} \dots \dots \dots (2)$$

حيث أن w معامل الوزن المعطل وتكون قيمته  $c_1, c_2, [0, 1]$ : ثوابت التسارع وتكون قيمته بين  $r_1, r_2, [0, 2]$  الرقم العشوائي وتكون قيمته بين  $p\_best, [0, 1]$  يشير إلى أفضل موقع للجسيم الحالي الذي تم الحصول عليه من  $g\_best$ ، يشير إلى أفضل بين جميع الجسيمات  $p\_best$ .

بالإضافة تكون حدود الموقع بين  $[x_{min}, x_{max}]$ ، اما حدود السرعة  $[v_{min}, v_{max}]$ ، يمكن تحديد خطوات عمل خوارزمية كما يلي:

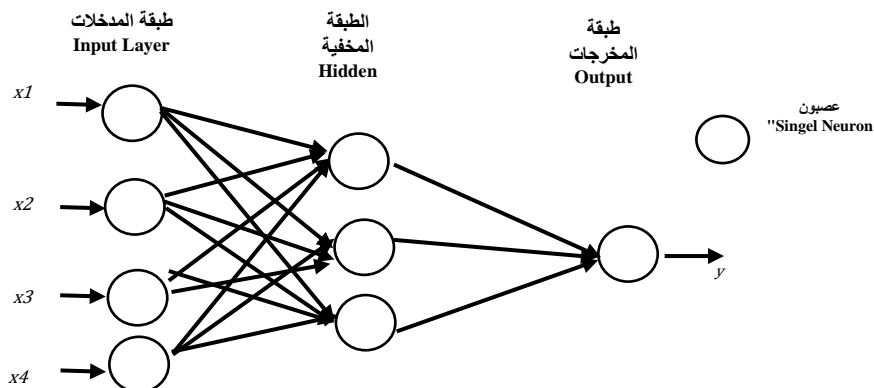
- 1- تحديد قيم أولية لعدد الجسيمات وموقع وسرعة لكل جسيم بشكل عشوائي.
- 2- تقييم قيمة اللياقة لكل الجسيمات في السرب بالاعتماد على دالة اللياقة والغرض من مشكلة التحسين.
- 3- المقارنة بين قيم اللياقة لكل جسيم في السرب وقيمة (pbest)، فإذا كانت القيمة أفضل من (pbest) يتم احلالها بدلا منه عكس ذلك تبقى قيم (pbest) كما هي.
- 4- تحديد أفضل قيمة لياقة بين كل الجسيمات في السرب، ومن ثم مقارنة قيمة اللياقة مع (gbest). فإذا كانت قيمة اللياقة أفضل من (gbset) الحالية تحل محله وإذا كانت قيمة (gbset) الحالية أفضل تبقى كما هي.
- 5- تحديث الموقع والسرعة لكل الجسيمات من طريق العلاقة (1)(2).
- 6- التوقف إذا حصلنا على أفضل قيمة لياقة (الحل الأمثل) اقل نسبة خطأ، او اذا توصلنا إلى أعلى حد من التكرار، عكس ذلك نعيد عمل الخطوات من (2 إلى 5) كما في الشكل (1-2).



طبقت خوارزمية الجسيمات (PSO) بنجاح في العديد من المجالات كخوارزمية تحسين ، تدريب الشبكات العصبية الاصطناعية، التعرف على الانماط (pattern recognition) ،انظمة السيطرة الضبابية (fuzzy system control) كونها سهلة الاستخدام وبسيطة الفهم والتطبيق ، لديها امكانية عالية الى الاستكشاف ولا تحتاج الى تعديل المعلمات .

3- الشبكات العصبية (Artificial Neural Network (ANN): نموذج رياضي يحاكي في عمله الدماغ البشري للانسان ومستوحاة من عمل الخلايا العصبية البيولوجية للدماغ البشري [11] ، تستطيع الخلايا العصبية التكيف والتعلم نتيجة كما هو الحال للخلاية العصبية ، مما يسمح لها حل المشكلات دون تدخل بشري والتي لا تحتاج معرفة مسبقة بهيكلية البيانات . تتكون الشبكة العصبية من ثلاثة اجزاء [12]، طبقة المدخلات (Input layer) ، والطبقات المخفية (hidden layer) وطبقة الإخراج (output layer) وكما في الشكل (2-2). تمثل طبقات الإدخال سلسلة من الخلايا العصبية التي تحتوي على إشارة الإدخال التي سيتم نقلها إلى الطبقات المخفية، تمثل هذه الطبقات قلب الشبكة العصبية يتم فيه تسليط الضوء على العلاقات بين المتغيرات المختلفة. النتيجة النهائية، في كثير من الأحيان التنبؤ النتيجة هي في طبقة الإخراج [7].

الشكل (2-2) شبكة عصبونية إصطناعية [13].



3-1 التعلم العميق (Deep Learning):

جزء فرعي من التعلم الآلي (Machine Learning)، ويشابه في هيكلته ووظائفه وأفعاله الدماغ البشري، وهو حقل متخصص ظهر مؤخراً في بدايات القرن الواحد والعشرين، يتم بناءه باستخدام الشبكات العصبية يحتوي على عقد قرار واعدة، وتعتبر الشبكة عميقة إذا تم تمرير البيانات المدخلة إلى سلسلة من التحويلات الغير خطية قبل استخراج النتائج [7].

### 2-3 الشبكات العصبية المتكررة (RNN) Recurrent Neural Network:

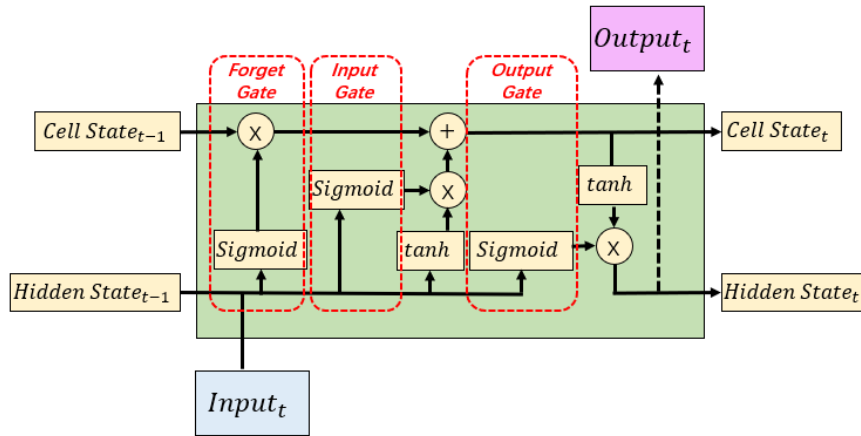
نوع من أنواع الشبكات العصبية تم استخدامها لأول مرة في 1980 [14]، لحل مشكلات شبكات العصبية الاصطناعية (ANN)، حيث تحتوي على تكرار (Recurrent) داخل الشبكة مما يعطيها عمليات تأثير الذاكرة من خلال استخدام بنية قائمة على حالة الذاكرة (Memory State) الذي يستخدم المعلومات السابقة في تقدير القيمة التي تليها وهذه الشبكة مفيدة جداً للتعرف على التسلسل (Sequence Dependent) بيانات معينة للتنبؤ بالقيمة الآتية للسلسلة حيث تحتفظ ببعض الميزات المهمة للسلسلة، كذلك التوقع والتنبؤ في البيانات المستقبلية وتعاني هذا النوع من الشبكات من مشكلة تلاشي وتضخم الاشتقاق التدريجي

(Exploding –Vanishing gradient problem) [15] [16].

### 3-3 الذاكرة الطويلة قصيرة المدى (LSTM) Long Short Term Memory:

تعتبر (LSTM) نوع خاص من (RNN)، قدمت لأول مرة من العالمين الالمانيين (Hochreiter & Sochmidhuber) في عام 1997 ولها القابلية على التعليم طويل المدى بتوظيف حالة الخلية (Cell Stat)، حيث لديها القدرة على المحافظة على حال الخلية ومن ثم إرسالها الى الخلية التالية لغرض ربط البيانات السابقة بالبيانات الحالية وذلك لغرض عدم فقدان البيانات من جهة، وتصميم (LSTM) بشكل صريح لحل مشكلة تلاشي التدرج عند تحديث الأوزان (Update Weight) التي نصل بها الى أفضل توقعات للنموذج بالشبكة العصبية المتكررة [17]. إذ إن الذاكرة (Memory) التي تحدد المعلومات التي يجب تمريرها بنحو أكبر في كل مرحلة زمنية، وكما في الشكل (2-3) [16].

### الشكل (2-3) هيكلية الذاكرة طويلة قصيرة المدى (LSTM) [16].



تستخدم في حل مشكلة تلاشي وتضخم الاشتقاق التدريجي (Exploding –Vanishing gradient problem) في التدريب الموجودة في الشبكات العصبية المتكررة (RNN) تتكون خلية الذاكرة طويلة قصيرة المدى من خلية عصبية متكررة بسيطة (RNN Cell) وحالة الخلية (Cell State) بالإضافة الى ثلاث بوابات وهي بوابة الإدخال (Input Gate  $i_t$ ) و بوابة الإخراج (Output Gate  $o_t$ ) وبوابة النسيان (Forget Gate  $f_t$ ). ويتلخص عمل الذاكرة طويلة قصيرة المدى (LSTM) باستخدام العلاقات الرياضية التالية من (9) الى (14):

$$f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \quad \dots \dots \dots (9)$$

$$j_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \dots \dots \dots (10)$$

$$\check{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \dots \dots \dots (11)$$

$$C_t = f_t * (C_{T-1} + i_c * \check{C}_t) \dots \dots \dots (12)$$

$$o_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \dots \dots \dots (13)$$

$$h_t = O_t * \tanh(C_{t-1}) \dots \dots \dots (14)$$

حيث ان  $f_t$ : قيمة المخرجة لبوابة النسيان،  $\sigma$  تمثل دالة التنغيم او التنشيط.  $w_f$  الوزن النسبي للقيم في دالة النسيان،  $h_{t-1}$ : قيمة المخرجات عند دالة النسيان في التسلسل الزمني  $t-1$ ،  $x_t$  قيمة المدخلات عند التسلسل الزمني لبوابة النسيان،  $b_f$  التحيز عند بوابة النسيان،  $j_t$ : قيمة الادخل في التسلسل الزمني  $t$ ،  $\check{C}_t$ : يمثل القيمة الحالية للخلاية بعد التعديل في التسلسل الزمني  $t$ :  $\tanh$  دالة التنغيم.

4- معايير الدقة (Accuracy Measurement): - لغرض تقييم أداء النموذج، تم استخدام معايير الدقة الاحصائية التالية، متوسط النسبة المئوية للخطأ المطلق (MAPE)، الارتباط الخطي (R)، متوسط الخطأ المطلق (MAE) حدد في المعادلة التالية [18][4]:

$$R = \frac{\sum_{t=1}^N (y_t - \hat{y}_t)(\hat{y}_t - \hat{y}_t)}{\sum_{t=1}^N (y_t - \hat{y}_t)^2 (\hat{y}_t - \hat{y}_t)^2} \dots \dots \dots (15)$$

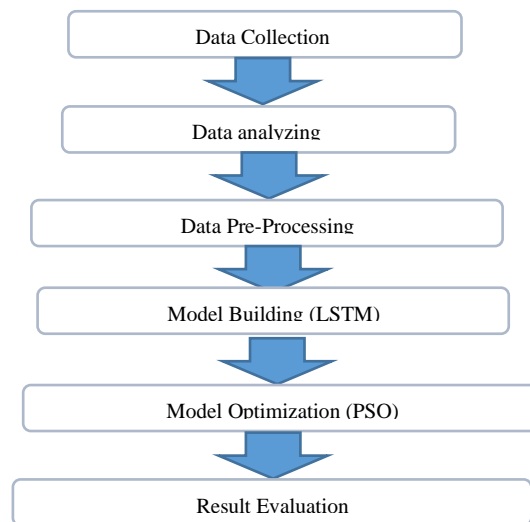
$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \dots \dots \dots (16)$$

$$MAPE = \frac{1}{N} \sum_{l=1}^N \frac{|Y_l - \hat{Y}_l|}{Y_l} \dots \dots \dots (17)$$

ثانياً: الجانب العملي :

في هذا الجانب من الدراسة سيتم تنفيذ خطوات الجانب العملي من الدراسة باستخدام برنامج التحليل الاحصائي (SPSS) لغرض اجراء اختبار الانحدار للبيانات بالإضافة الى لغة برمجة الذكاء الاصطناعي (Python) لغرض بناء إنموذج التنبؤ محل الدراسة (LSTM) وبناء خوارزمية التحسين (PSO) من ثم المقارنة بين النماذج بواسطة معايير الدقة التي تم ذكرها في الجانب النظري من الدراسة ، والشكل (3-1) يوضح خطوات الجانب العملي من الدراسة .

الشكل (3-1) يوضح خطوات الدراسة





1- **تجميع البيانات Data Collection** :- تم تجميع بيانات الدراسة من الموقع الرسمي بالبنك المركزي ووزارة التخطيط والموقع الرسمي للبنك الدولي للفترة من 2003 إلى 2022 وبشكل شهري، (إيرادات النفط، إيرادات الضرائب، الإيرادات الأخرى، سعر الصرف، حجم الدين العام، سعر برميل النفط، الاستثمار الأجنبي، الاستثمار الداخلي، الزيادة السكانية، مؤشر سوق الأوراق المالية العراقي، الدخل القومي، الفائض والعجز، الإيرادات الفعلية). تم جمع البيانات يدوياً وعلى نموذج ورقي، تم تحويلها إلى صيغة إلكترونية على شكل جداول (Excel) لكي يسهل التعامل معها لاحقاً.

2- **تحليل البيانات (Data Analysis)** : تم تحليل البيانات باستخدام برنامج التحليل الاحصائي (SPSS) لمعرفة معنوية للمتغير التابع الإيرادات الضريبية والمتغيرات التفسيرية، لكون قيمة (Sig F) اقل من القيمة الجدولية عند مستوى دلالة ( $\alpha=0.05$ ) كما نلاحظ أن قيمة معامل التحديد ( $R^2$ ) تساوي (0.819) وهذا ما يبين أن المتغيرات التفسيرية التي يتضمنها النموذج تفسر (89.1%) من التغيرات التي يتأثر بها المتغير التابع والجدول (1-3) يبين معنوية الانحدار.

الجدول (1-3) معنوية انحدار

المتغير التابع	R	$R^2$	F	Sig F
الإيرادات الضريبية	0.905	0.819	71.749	0.000
المتغيرات التفسيرية	Beta	T	Sig t	
الإيرادات الأخرى	0.138	2.687	0.008	
الإيرادات النفطية	-0.389	-1.911	0.057	
الفائض والعجز	0.079	0.980	0.328	
النفقات	1.046	5.598	0.000	
الاصول الثابتة	-0.258	-1.378	0.170	
سعر برميل النفط	0.208	4.024	0.000	
الدين العام	0.843	6.711	0.000	
الاستثمار الاجنبي	-0.090	-1.428	0.155	
الاستثمار العراقي	-0.073	-0.781	0.436	
سعر الصرف	-0.387	-5.070	0.000	
الزيادة السكانية	0.130	0.875	0.383	
الدخل الوطني	-0.518	-2.321	0.021	
مؤشر سوق العراق	0.10	0.321	0.748	

3- **المعالجة المسبقة (Data Pre-Processing)** :- تنقسم مرحلة المعالجة المسبقة للبيانات إلى مرحلتين تقسيم البيانات (Data Splitting) وتقييس البيانات (Features Scaling) إذ في هذه المرحلة يتم تقسيم البيانات الى ثلاث مجموعات مجموعة تدريب (Training Set) ومجموعة التحقق (Validation Set) ومجموعة الإختبار (Testing Set) (80%) تدريب (10%) تحقق (10%) إختبار، ثم إجراء التقييس للبيانات وجعلها بين الصفر والواحد باستخدام (Min-Max) لجعلها ضمن حدود دالة التفعيل الذي تستثار عنده وكما في العلاقة (18)[19].

$$\dots\dots\dots(18)$$

$$x_{scal} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

حيث تمثل x القيمة الفعلية في مجموعة بيانات السلاسل الزمنية، ويشير x\_min إلى الحد الأدنى للقيمة في مجموعة البيانات الفعلية، ويمثل x-max الحد الأقصى للقيمة في مجموعة البيانات، ويقدم x\_scal القيمة الناتجة بعد إجراء التقييس وتكون بين [0-1].

4- بناء إنموذج التنبؤ الذاكرة طويلة فصيصة المدى (LSTM):- بإستخدام لغة بايثون ومن مكتبة (Tensor flow) و(Keras) تم بناء الإنموذج بواسطة ثلاثة طبقات مخفية (Hidden Layer) بالإضافة الى طبقتي المدخلات (Input Layer) والمخرجات (Output Layer)، و بإستخدام دالة التفعيل Relu Activation Function [19]. في الطبقات المخفية (Hidden Layer) كذلك دالة التفعيل (Linear Activation Function) في طبقة الإخراج (Output Layer)، وتم تدريب الإنموذج بإستخدام خوارزمية الإنتشار العكسي (BP) وبمعدل تعلم (Learning Rate=0.0001) وإستعمال المحسن (Adam) لغرض تحديث الازان (Weight) والتحيز (Basie) في كل دورة (epoch) مكرراً 1000 دورة لحين الوصول الى أقل خطأ ممكن.

5- تحسين اداء الانموذج باستخدام خوارزمية التحسين الجسيمات (PSO):- بعد بناء النماذج يتم تحسين عمل أنموذج التنبؤ باستخدام خوارزمية الجسيمات (PSO) ، تتوافق خوارزمية (PSO) مع خوارزميات البحث العشوائية إذ يجب تحديد عدد افراد السرب (Population Size) ،أقصى عدد للدورات ،وتحديد قيم الثوابت (C2,C1) كذلك تحديد الازان الابتدائية (W) ، بالإضافة تحديد مجال معين للبحث عن المعلمات المراد تحسينها و إن أفضل تحسين للانموذج هو (Bach ,Dropout) (Neuron,size) ، اذ تم تحديد الحدود الدنيا والعليا لكل معلمة لغرض تحديد مجال البحث لكل معلمة تحسين والتي تضمن ان يتم البحث عن معلمة التحسين ضمن النطاق المحدد لها ، تم تمثيل هيكلية الشبكة العصبية من طريق كل فرد من افراد السرب مع المعلمات المرشحة لغرض تهيئة سرب الجسيمات لخوارزمية التحسين (PSO) ، يشير كل مجال تحسين فردية للشبكة ويتم انشاؤها بواسطة قيم عشوائية حقيقية ضمن مجال البحث المخصص ، بعد عمليات البحث نحصل على مجموعة من هيكليات الشبكة المقترحة ، ويمكن تقييم المعلمات المختارة بتدريب الشبكة وفق المعلمات المرشحة والتحقق من صحتها والجدول (2-3) ، يوضح المعلمات المستخدمة لخوارزمية الجسيمات (PSO).

جدول (2-3) معلمات خوارزمية الجسيمات (PSO)

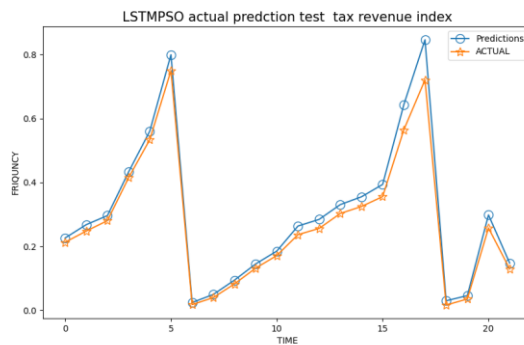
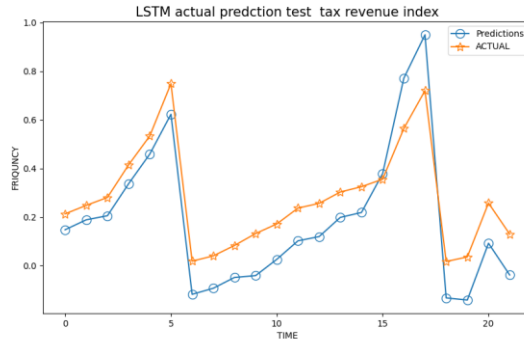
المعلمات	القيمة	الوصف
EPOCH	1000	عدد الدورات
Population Size	4	عدد افراد السرب
C1	1.429	ويسمى الازان المعرفة ويحدد تأثير افضل موضع للجسيم
C2	1.429	الازان الاجتماعية ويحدد افضل موقع للجسيم العالمي ويتم اقتراح القيم من خلال البحث في خوارزمية الجسيمات
W	0.8	يحدد تأثير السرعة الحالية
r 1 and r 2	0 -1	القيم التي تضيف تأثير عشوائي للتأثير في حركة الجسيمات

6-تقييم النتائج (Result Evaluation):- في هذه المرحلة يتم تقييم النتائج المستخرجة من كل إنموذج بإستخدام معايير الدقة

(R, MAPE,MAE) لقياس دقة التنبؤ للإنموذج المقترح ، وإختيار الأفضل منها لإعتمادها من متخذي القرار والمهتمين. ولغرض التأكد من كفاءة عمل الأنموذج الذي تم بناؤه تم مقارنة بيانات الإخراج مع البيانات الحقيقية بالأعتماد على معايير الدقة الاحصائية لمعرفة مدى تقارب البيانات الحقيقية والبيانات المتوقعة والشكل (2-3) و (3-3) يوضح تقارب كبير مع البيانات الحقيقية ، مما يؤكد كفاءة التدريب.



## الشكل (2-3) القيم الفعلية والقيم التنبؤية باستخدام انموذج (LSTM).



## (3-3) القيم الفعلية والقيم التنبؤية (LSTMPSO).

الشكل (3-3) يوضح تقارب القيم الفعلية والقيم الحقيقية باستخدام الإنموذج (LSTMPSO) في التنبؤ بكل من الإيرادات الضريبية حيث يمثل اللون الأزرق القيم المتوقعة ويمثل اللون البرتقالي القيم الفعلية للسلاسل الزمنية، بينما الشكل (2-3) تقارب أقل دقة بين القيم الفعلية والقيم التنبؤية باستخدام الإنموذج (LSTM) منفرداً بدون خوارزمية التحسين الجسيمات.

## جدول (3-3) معايير الدقة للتنبؤ بالإيرادات الضريبية

Model	MAPE	R	RMSE
LSTMPSO	0.159828	0.997382	0.00837
LSTM	0.560547	0.961091	0.01316

يبين الجدول (3-3) مقاييس الدقة لغرض التنبؤ بالإيرادات الضريبية باستخدام أنموذج (LSTM) و (LSTMPSO)، ويتضح من معايير الدقة أن الأنموذج المقترح (LSTMPSO) أكثر دقة من (LSTM) في التنبؤ بالإيرادات الضريبية وذلك بمتوسط خطأ نسبي مطلق (0.159828) MAPE، ومعامل ارتباط R (0.997382).

## ثالثاً: الاستنتاجات والتوصيات :

## أ- الاستنتاجات:

- 1- استخدام أسلوب علمي حديث بالتنبؤ بالإيرادات الضريبية وهو (PSO) و (LSTM)، وذلك لغرض تلافي العيوب الناتجة من استخدام الأساليب السابقة.
- 2- تعتبر (LSTM) واحدة من أهم التقنيات الديناميكية التي لها القابلية على تحليل البيانات الغير خطية ولها القدرة على التنبؤ الدقيق وذلك لقدرتها على التحليل العميق للعلاقات بين البيانات.
- 3- ان تخمين الإيرادات الضريبية بالاعتماد على الأساليب الشخصية والتقليدية دون الاعتماد على الاسس العلمية يؤثر سلباً على الإيرادات لأن التنبؤ التدقيق يساعد في التشخيص المبكر.



4- تعتبر خوارزمية التحسين الجسيمات (PSO) من الأساليب المهمة التي تساهم في زيادة دقة التنبؤ كما في الجداول (3-4) وذلك لقدرته العالية على اختيار افضل معاملات لانموذج التدريب.

#### ب- التوصيات:

- 1- يوصي الباحث تطبيق تقنيات حديثة وادخال تكنولوجيا المعلومات عند تخمين الإيرادات الضريبية .
- 2- ضرورة الاعتماد على الاسس العلمية عند تخمين التدفقات النقدية الضريبية من دون الاعتماد على اسلوب التحكيم الشخصي.
- 3- توفير البيانات من قبل الجهات ذات العلاقة وذلك لتحقيق المنفعة المتبادلة بين الباحثين والجهات ذات العلاقة.
- 4- استخدام الاساليب الذكية وخوارزميات تحسين اخرى ومقارنة النتائج.

#### References:

- 1- S. abd alhalim and H. Mahmood, "The relationship between federal budget estimates and actual implementation," J. Account. Financ. Stud., vol. 13, no. 43, pp. 100–120, Jun. 2018, doi: 10.34093/jafs.v13i43.56.
- 2- C. Technology and M. Dissertations, "Prediction of annual revenue collection by using data mining techniques : a case of local government authorities in Tanzania," p. 85, 2020, [Online]. Available: <http://hdl.handle.net/20.500.12661/2697>
- 3- C. Chimilila, "Forecasting Tax Revenue and its Volatility in Tanzania," African J. Econ. Rev., vol. 5, no. 1, pp. 84–109, 2017, doi: 04/ag.econ.264561
- 4- K. IBTESAM, "(Dynamic& Static) Forecast of surplus or Deficit of Public budget in Iraq for (2017•2018)," J. Econ. Adm. Sci., vol. 25, no. 112, pp. 391–413, Jun. 2019, doi: 10.33095/j eas.v25i112.1670.
- 5- S. Hamidii and A. Hamid, "Forecasting the Iranian Tax Revenues : Application of Nonlinear Models," pp. 1–22, 2019.
- 6- B. Kozicki, T. Waściński, and A. Lisowska, "Methods of Forecasting Budget Revenues," Zesz. Nauk. Uniw. Przyr. w Siedlcach. Ser. Adm. i Zarządzanie, vol. 51, no. 51, pp. 23–32, Aug. 2020, doi: 10.34739/zn.2020.51.03.
- 7- M. Rhanoui, S. Yousfi, M. Mikram, and H. Merizak, "Forecasting financial budget time series: Arima random walk vs lstm neural network," IAES Int. J. Artif. Intell., vol. 8, no. 4, pp. 317–327, 2019, doi: 10.11591/ijai.v8.i4.pp317-327.
- 8- G. Kumar, U. P. Singh, and S. Jain, "An adaptive particle swarm optimization-based hybrid long short-term memory model for stock price time series forecasting," Soft Comput., vol. 26, no. 22, pp. 12115–12135, 2022, doi: 10.1007/s00500-022-07451-8.
- 9- Z. Dan, "Improving the accuracy in software effort estimation: Using artificial neural network model based on particle swarm optimization," in Proceedings of 2013 IEEE International Conference on Service Operations and Logistics, and Informatics, 2013, pp. 180–185. doi: 10.1109/SOLI.2013.6611406.
- 10- A. Anand and L. Suganthi, "Hybrid GA-PSO optimization of Artificial Neural Network for forecasting electricity demand," Energies, vol. 11, no. 4, 2018, doi: 10.3390/en11040728.
- 11- E. A. A. Hamra, "Combine Genetic Algorithm and Particle Swarm Optimization Approach for Neural Network Classification," Thesis, 2016.
- 12- I. Khandelwal, R. Adhikari, and G. Verma, "Time series forecasting using hybrid arima and ann models based on DWT Decomposition," Procedia Comput. Sci., vol. 48, no. C, pp. 173–179, 2015, doi: 10.1016/j.procs.2015.04.167.
- 13- Rusul Mustafa Ismaiel and Waleed Abdullah Araheemah, "Compare Some Classification Methods for COVID-19 Identification with Application," J. Tech., vol. 4, no. 4, pp. 228–236, Dec. 2022, doi: 10.51173/jt.v4i4.614.
- 14- S. A. Ludwig, "Comparison of Time Series Approaches applied to Greenhouse Gas Analysis: ANFIS, RNN, and LSTM," IEEE Int. Conf. Fuzzy Syst., vol. 2019-June, 2019, doi: 10.1109/FUZZ-IEEE.2019.8859013.
- 15- X. Xing, Z. Li, T. Xu, L. Shu, B. Hu, and X. Xu, "SAE+LSTM: A new framework for emotion recognition from multi-channel EEG," Front. Neurobot., vol. 13, no. June, pp. 1–14, 2019, doi: 10.3389/fnbot.2019.00037.
- 16- Z. Pan, "Advisor :," Illinois Urbana-Champaign, 2021.
- 17- Q. Tang, T. Fan, R. Shi, J. Huang, and Y. Ma, "Prediction of financial time series using LSTM and data denoising methods," arXiv Prepr. arXiv2103.03505, 2021, [Online]. Available: <https://arxiv.org/abs/2103.03505>
- 18- W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," PLoS One, vol. 12, no. 7, pp. 1–24, 2017, doi: 10.1371/journal.pone.0180944.
- 19- Sama Hayder Abdulhussein AlHakeem, Nashaat Jasim Al-Anber, and Hayfaa Abdulzahra Atee, "Iraqi Stock Market Prediction Using Artificial Neural Network and Long Short-Term Memory," J. Tech., vol. 5, no. 1, pp. 9–15, Mar. 2023, doi: 10.51173/jt.v5i1.846.