مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٨)العدد (٨) ٢٠٢٤

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (8) No. (8) 2024

تحسين استثمارات الأصول العقارية باستخدام الذكاء الاصطناعي لزيادة دقة التنبؤ بقيمة العقارات

دعاء مهنا *

(تاريخ الإيداع ٢٠٢٤/٣/٢٨ . قُبِل للنشر في ٢٠٢٤/٧/١٨)

□ ملخّص □

تصنف الأصول العقارية كواحدة من أهم أصناف الاستثمار التي يتم الاعتماد عليها كمصدر للدخل والثروة وتشمل الأصول العقارية مجموعة متنوعة من الممتلكات مثل المباني السكنية والتجارية والأراضي والمنازل والمكاتب والمحلات التجارية بالإضافة الى امتلاك الأصول العقارية خصائص مميزة تجعلها محل اهتمام كبير للمستثمرين، مثل استقرار القيمة على المدى الطويل، والقدرة على الاستفادة من تغيرات سوق العقارات لزيادة قيمتها ويمكن تحسين الاستثمار في الأصول العقارية وزيادة كفاءة عمليات اتخاذ القرارات بالاعتماد على تحليلات البيانات واستخدام النماذج العلمية.

يلعب الذكاء الاصطناعي دورًا مهمًا في تطور الأصول العقارية بفضل قدرته على تحليل ومعالجة كميات كبيرة من البيانات بسرعة ودقة مما أدى الى تحسين تقدير قيمة العقارات وقدرة المستثمرين على اتخاذ قرارات استثمارية مستندة إلى توقعات دقيقة لسوق العقارات.

يقدم البحث منهجية علمية للتنبؤ بقيمة العقار بشكل تلقائي بالاعتماد على مجموعة من بارامترات الدخل وقد اعتمدت هذه المنهجية على:

- ✓ عمليات المعالجة المسبقة للبيانات لتصبح مهيئة للاستخدام ضمن نموذج الذكاء الاصطناعي.
- ✓ استخدام الشبكة العصبونية الاصطناعية للتنبؤ بقيمة العقار.
- ✓ تحسين دقة الشبكة العصبونية.
- ✓ تصميم واجهة تفاعلية للعميل والمستثمر.

الكلمات المفتاحية: الأصول العقارية، الذكاء الاصطناعي الشبكات العصبونية الاصطناعية، التنبؤ للعقارات، عمليات معالجة البيانات

*ماجستير في قسم تكنولوجيا المعلومات–كلية تكنولوجيا المعلومات والاتصالات–جامعة طرطوس–طرطوس–سوريا

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٨)العدد (٨) ٢٠٢٤

Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (8) No. (8) 2024

Enhancing Real Estate Investments Using Artificial Intelligence to Improve Accuracy in Property Value Prediction

D'uaa mhnaa*

(Received 28/3/2024 . Accepted 18/7/2024)

□ ABSTRACT □

Real estate assets are considered one of the most significant investment types relied upon as a source of income and wealth. They encompass a diverse range of properties such as residential and commercial buildings, lands, houses, offices, and shops. Real estate assets possess unique features that make them highly attractive to investors, such as long-term value stability and the ability to benefit from real estate market changes to increase their value. Enhancing real estate investments and improving decision-making efficiency can be achieved by utilizing data analytics and scientific models.

Artificial intelligence plays a crucial role in the evolution of real estate assets due to its capability to analyze and process large volumes of data quickly and accurately. This has led to the improvement of property valuation estimation and investors' ability to make investment decisions based on precise real estate market forecasts. The research provides a scientific methodology for automatically predicting property values based on a set of income parameters. This methodology relies on:

- ✓ Data preprocessing operations to make them suitable for use in artificial intelligence models.
- ✓ Using artificial neural networks to predict property values.
- ✓ Enhancing the accuracy of the neural network.
- ✓ Designing an interactive interface for clients and investors.

Keywords: Real estate assets, Artificial intelligence, Artificial neural networks, Property prediction, Data preprocessing operations**.**

*Postgraduate و Department of Information Technology, Faculty of Information and Communication Technology, University of Tartous, Syria

١ – المقدمة

تلعب العقارات دورًا حاسمًا في الاقتصاد العالمي في ظل التطور السريع الذي تشهده الاقتصادات العالمية وتنوع الاستثمارات العقارية بحيث تعتبر العقارات من أهم الأصول الاقتصادية التي تجذب اهتمام المستثمرين والمتخصصين بالإضافة الى كونها تعتبر ملاذًا آمنًا للاستثمار بل تعتبر أيضًا عاملًا محفزًا للتنمية الاقتصادية والاجتماعية والثقافية كما شكلت العقارات مصدر رئيسي للدخل للكثير من الأفراد والشركات وحكومات الدول، وأسهمت في توفير فرص العمل وتعزيز الاستقرار الاقتصادي

ارتبطت تقنيات الذكاء الاصطناعي بقطاع العقارات بشكل وثيق مما أدى الى تحسين تقديرات القيمة العقارية، وتحليل البيانات العقارية، وتطوير نماذج التنبؤ بالأسعار العقارية، وتحسين عمليات التسويق وإدارة الممتلكات من خلال الاعتماد على الشبكات العصبية الاصطناعية والخوارزميات الوراثية وخوارزميات التعلم الالي.

يتم استخدام الذكاء الاصطناعي كدعم تكنولوجي للمستثمرين في مجال العقارات وتعد مفيدة لعدة أسباب:

- تحلل خوارزميات الذكاء الاصطناعي كمية كبيرة من البيانات للتنبؤ بالاتجاهات المستقبلية في أسعار العقارات والطلب.
- تقييم دقيق لقيمة العقارات من خلال الاخذ بعين الاعتبار عدة عوامل (الموقع والحجم والمرافق واتجاهات السوق).
- تقديم توصيات عقارية شخصية بناءً على تفضيلات ومتطلبات الأفراد.

تم استخدم تقنية تنقيب البيانات بشكل كبير ضمن العقارات ويعود ذلك الى قدرتها على استخراج المعلومات ذات الأهمية من البيانات الخام، مما يجعلها ذات فائدة كبيرة في توقع أسعار المنازل والميزات الرئيسية للسكن، كما أوضحت الدراسات أن اختلاف أسعار العقارات يمكن أن تؤثر بشكل كبير على أصحاب المنازل وسوق العقارات وعليه تم استخدام نماذج لتوقع أسعار العقارات بشكل يؤدي الى الحصول على أكبر دقة ممكنة ومن هذه النماذج:

- ✓ الشبكة العصبية الاصطناعية.
- ✓ متجهات الدعم (SVR).
- ✓ XGBoost

أكدت النتائج على فعالية استخدام النماذج السابقة للتنبؤ بأسعار العقارات مع الاخذ بعين الاعتبار العوامل المكانية ووكلاء العقارات[1] .

تكمن الفائدة من استخدام تقنيات الذكاء الاصطناعي ضمن مجال العقارات من خلال التقييم التلقائي للأسعار وتوجد العديد من الطرق لتحقيق ذلك منها نماذج الانحدار المتعدد والتعلم الآلي، وتشير الدراسة الى الصعوبات التي تواجه أساليب التعلم الآلي الحديثة المتمثلة في نقص المعرفة بتلك الأساليب، مما يؤدي إلى قلة الوعي بفاعليتها في عملية تقدير قيمة.

أجرت الدراسة مقارنة بين نموذجي انحدار متعدد والشبكات العصبونية لتحديد أسعار المنازل باستخدام نفس مجموعة البيانات. أظهر النموذج الانحداري التعددي نتائج متوسطة وكان ذات فعالية عالية لأغراض التحليل، بينما لم تحقق الشبكات العصبية أداء متميزاً، وتم تقدير الأسعار بشكل مبالغ فيه في معظم الحالات، لكن يعود ذلك إلى أهمية تعزيز البحث والتعلم لتحقيق نجاح أكبر في استخدام الشبكات العصبية مع مجموعات بيانات أكبر[2] .

تجدر الإشارة الى أنه تم تطبيق نماذج للتنبؤ بأسعار العقارات باستخدام طرق التعلم الآلي بشكل فعلي ضمن محافظة أنقرة(تركيا)بالاعتماد على مجموعة بيانات مبيعات العقارات في يونيو ويوليو ٢٠٢١ حيث تم تضمين بيانات العقارات الحقيقية التي تم الحصول عليها عبر الإنترنت في عملية تحضير البيانات المفصلة أولاً ثم استخدام مجموعة من تقنيات العلم الالي (Lasso and Ridge Regression, XGBoost and Artificial Neural Networks (ANN)) وفقًا للنتائج التجريبية، يظهر أن XGBoost وشبكات العصبونية الصناعيةِ(ANN) تعتبر تقنيات واعدة في توقع أسعار مبيعات العقارات[3].

في هذا البحث تم العمل على:

- ✓ تحميل قاعدة البيانات.
- ✓ تنفيذ عمليات المعالجة المسبقة على قاعدة البيانات لتهيئتها لعمليات التدريب.
- ✓ بناء نموذج التنبؤ بالاعتماد على الشبكة العصبونية الاصطناعية.
- ✓ تحسين نتائج النموذج المقترح.
- ✓ توثيق النتائج النهائية.

2- أهمية البحث وأهدافه

تشهد الاقتصادات العالمية تطوراً متسارعاً، وتعتبر العقارات من القطاعات الحيوية والحاسمة التي تسهم في تشكيل الركيزة الأساسية للاقتصاد العالمي حيث تلعب دورًا مهمًا للغاية في تعزيز الاستقرار الاقتصادي والاجتماعي، وتسهم في تعزيز الثقة بالسوق وجذب الاستثمارات، ومن الجدير بالذكر أن العقارات تُعد من أهم الأصول الاقتصادية التي تجذب اهتمام المستثمرين بسبب استقرار قيمتها على المدى الطويل وقدرتها على توفير العوائد المالية المستدامة.

أدت تقنيات الذكاء الاصطناعي الى تقدم ملحوظ في مجال العقارات وذلك يعود الى قدرتها على تحليل البيانات الكبيرة والمعقدة بشكل يؤدي الى توقعات أكثر دقة لأسعار العقارات وبموثوقية عالية بالإضافة الى المساهمة في توجيه المستثمرين نحو الفرص الاستثمارية الصحيحة وتقليل المخاطر المحتملة.

وبناءً على ما تقدم فإن الهدف من البحث هو بناء نموذج ذكاء اصطناعي قادر على التنبؤ بأسعار العقارات بأكثر دقة ممكنة متضمناً عمليات المعالجة المسبقة التي ستطلبها البيانات لتكون جاهزة للاستخدام ضمن هذا النموذج

3- طرائق البحث ومواده

تم العمل ضمن هذا البحث على بناء نموذج ذكاء اصطناعي قادر على التنبؤ بأسعار العقارات بأعلى دقة ممكنة وفق التالي:

- تهيئة البيانات المستخدمة لتكون جاهزة للاستخدام في النموذج.
- تقسيم البيانات إلى مجموعتي تدريب واختبار.
- بناء الشبكة العصبونية الاصطناعية(ANN) بواجهة مخصصة لتوقع أسعار العقارات تتضمن تحديد عدد الطبقات المخفية وعدد العصبونات وتابع التنشيط ومعدل التعلم وعدد مرات التدريب.
- تحسين وتوثيق النتائج.

هذا المنهج نُفذ باستخدام المحاكاة الحاسوبية بالاعتماد على لغة Python (3.11.5) من خلال برنامج Spyder (٥,٤,٣)وباستخدام قاعدة البيانات التي تم تحميلها من موقع kaggel من خلال الرابط

التالي-[4]https://www.kaggle.com/datasets/yasserh/housing-prices
dataset?resource=download والتي تتضمن معلومات عن :

- ✓ السعر (Price)
- ✓ المساحة (Area)
- ✓ عدد غرف النوم (Bedrooms)
- ✓ عدد الحمامات (Bathrooms)
- ✓ عدد الطوابق (Stories)
- ✓ الطريق الرئيسي (Mainroad)
- ✓ غرفة الضيوف (Guestroom)
- ✓ الطابق السفلي من المبنى (Basement)
- ✓ تدفئة باستخدام المياه الساخنة (Hotwaterheating)
- ✓ التكييف(Airconditioning)
- ✓ موقف سيارات (Parking)
- ✓ المنطقة المفضلة (Prefarea)
- ✓ حالة الأثاث(Furnishingstatus)

3-1 معالجة البيانات:

تعتبر المعالجة المسبقة للبيانات (Data preprocessing) عملية يتم فيها تنظيف البيانات وتحويلها وإعدادها قبل استخدامها ضمن نموذج، تحتوي البيانات الخام(Raw data)على معلومات غير ذات صلة(irrelevant)أو زائدة عن الحاجredundant)(أو قيم مفقودة(missing values) أو تنسيقات غير متسقة(inconsistent formats) يمكن أن تؤثر سلبًا على دقة وموثوقية نماذج ذكاء اصطناعي لذلك، تعد المعالجة المسبقة للبيانات خطوة ضرورية لأنها تساعد على ضمان أن تكون البيانات بالتنسيق والجودة المناسبين للنموذج للتعلم بفعالية[5] .

٣-١-١ أنواع تقنيات المعالجة المسبقة للبيانات:

تصنيف تقنيات المعالجة المسبقة للبيانات[5] إلى فئتين:

- ➢ **تنظيف البيانات(data cleaning):** تعد خطوة مهمة جداً في المعالجة المسبقة للبيانات التي تتضمن تحديد وتصحيح أو إزالة الأخطاء والتناقضات في مجموعة البيانات وذلك يعود الى أن مجموعات البيانات الواقعية غالبًا ما تكون غير كاملة أو صاخبة أو تحتوي على معلومات غير ذات صلة أو زائدة عن الحاجة وتتضمن:
- ✓ **معالجة البيانات المفقودة(Handling Missing Data):** تؤثر البيانات المفقودة بشكل كبير على دقة وموثوقية نماذج التعلم الآلي، وتوجد العديد من الأساليب للتعامل مع البيانات المفقودة منها:
  - **التضمين(imputation)**تتضمن استبدال القيم المفقودة بقيم منطقية بناءً على البيانات المتاحة
  - **الحذف (deletion)**تتضمن إزالة البيانات التي تحتوي على قيم مفقودة.
  - **التقدير(estimation)**تتضمن استخدام النماذج الإحصائية للتنبؤ بالقيم المفقودة بناءً على البيانات المتاحة.

○**اكتشاف وإزالة القيم المتطرفة (Outlier Detection and Removal)**يتضمن تحديد وإزالة البيانات التي تختلف اختلافًا كبيرًا عن بقية مجموعة البيانات والتي قد توثر سلبًا على دقة وموثوقية من خلال تحريف توزيع البيانات.

➢ **تحويل البيانات(data transformation)**:هو عملية تعديل البيانات بطريقة يمكنها تحسين دقة وكفاءة نماذج التعلم الآلي وتتضمن:

○**تحجيم الميزات (feature scaling)**: هو أسلوب شائع يستخدم في تحويل البيانات يتضمن تحجيم ميزات مجموعة البيانات إلى نطاق قياسي ويعود السبب في ذلك الى أن بعض خوارزميات التعلم الآلي حساسة لمقياس الميزات، ويمكن أن تؤدي إلى نتائج غير مرضية إذا لم يتم تسوية الميزات. هناك عدة تقنيات لتحجيم الميزات مثل تقنية min–max scaling التي تقوم بتحديد القيم الأدنى والأقصى، وتقنية z–score normalization التي تقوم بتسوية البيانات بوحدة انحراف معياري، وتقنية robust scaling التي تقوم بتحجيم البيانات بشكل قوي.

○**اختيار الميزة (feature selection)**:هو عملية تحديد مجموعة فرعية من الميزات الأصلية الأكثر صلة بالمشكلة التي يتم حلها و يمكن أن تساعد تقنيات اختيار الميزة في تقليل تعقيد النموذج وتحسين قدرته على التعميم، تتضمن بعض تقنيات تحديد الميزة الشائعة طرق التصفية filter methods وطرق الغلاف wrapper methods والطرق المضمنة.embedded methods

○**استخراج الميزات (feature extraction)** هو عملية إنشاء ميزات جديدة من الميزات الأصلية لمجموعة البيانات ويمكن أن يساعد ذلك في التقاط الأنماط والعلاقات المهمة في البيانات التي قد لا تكون واضحة من الميزات الأصلية و تشمل تقنيات استخراج الميزات تحليل المكونات الرئيسية(principal component analysis (PCA ، وتحليل المكونات المستقلة(independent component analysis (ICA) ، وعامل المصفوفة غير السلبي–non negative matrix factorization (NMF).

3–2 الذكاء الاصطناعي:

يُعتبر الذكاء الاصطناعي[6]أحد فروع علم الحاسوب، وإحدى الركائز الأساسية التي تقوم عليها تكنولوجيا الصناعة في العصر الحالي، ويُمكن تعريف مصطلح الذكاء الاصطناعي(Artificial intelligence)بأنه قدرة الآلات والحواسيب الرقميّة على القيام بمهام مُعينة تُحاكي وتُشابه تلك التي تقوم بها الكائنات الذكيّة؛ كالقدرة على التفكير أو التعلُم من التجارب السابقة أو غيرها من العمليات الأُخرى التي تتطلب عمليات ذهنية، كما يهدف الذكاء الاصطناعي إلى الوصول إلى أنظمة تتمتع بالذكاء وتتصرف على النحو الذي يتصرف به البشر من حيث التعلُم واتخاذ القرار بحيث تُقدم تلك الأنظمة لمستخدميها خدمات مختلفة من التعليم والإرشاد والتفاعل وما الى ذلك.

تشير أغلب تطورات تطبيقات الذكاء الصنعي إلى فئة من الخوارزميات المعروفة باسم تعليم الآلة (Machine Learning) ، وتَستخدم هذه الخوارزميات مفاهيمَ الإحصاء من أجل اكتشاف الأنماط عن طريق كميات هائلة من البيانات، ثم يتم استخدام تلك الأنماط في عملية التنبؤ للفئات.

وتعدُّ هذه الخوارزميات أساس العديد من التطورات المهمة، كالتعرف إلى الوجه، ولكن هذه الخوارزميات تعد جزء صغير ممّا يمكن أن يكون عليه الذكاء الاصطناعي.

إن تعلم الآلة والتعلم العميق مع وجود بيانات كافية سيؤدي في نهاية المطاف إلى الذكاء الصنعي العام

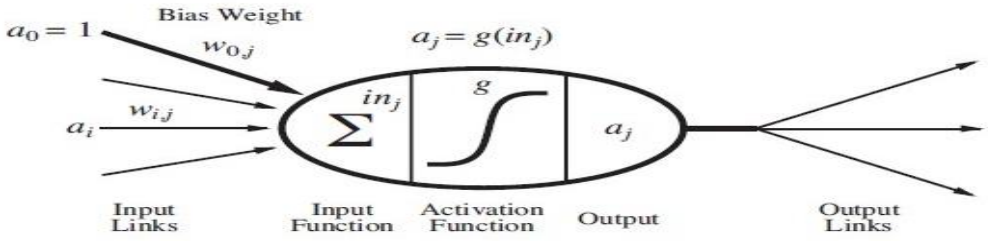
(Artificial General Intelligence (AGI)) ؛ أي أن تصبح الآلة قادرة على أداء أيّة مهمة يمكن للإنسان فعلها دون تدريب مسبق.

3-2 -١ الشبكات العصبونية:

تقنيات حسابية مصممة [7]لمحاكاة الطريقة التي يؤدي بها الدماغ البشري مهمة معينة من خلال معالجة ضخمة موزعة على التوازي بحيث تتشابه الشبكات العصبونية مع الدماغ البشري بقدرتها على اكتساب المعرفة بالتدريب، وتخزين هذه المعرفة باستخدام قوى وصل داخل العصبونات تسمى الأوزان(weights)

تتألّف الشبكات العصبونية الصنعية الموضحة في الشكل (١) من وحداتٍ مرتبطةٍ مع بعضِها البعض عبرَ وصلات (خط ارتباط)؛ هذه الوحدات هي عبارة عن عناصر حسابية تسمى العصبونات أو نيرونات (Neurons)، وتُستخدَم هذه الوصلات لنقل النّشاط بين هذه الوحدات، ويملكُ كلُّ خط ارتباط وزناً معيَّناً Weight يزداد بازدياد قوة الاتصال بين الوحدتين المرتبطتين عبر هذا الخط. تتوضَّعُ المعلومات التي نريد معالجتَها عند الطبقة الأولى من الوحدات، وقد يكون خَرْجُ كلّ عصبونٍ دَخْلاً لعصبونٍ آخر، كما تملكُ كلُّ وحدةٍ دَخْلاً وهميَّاً تساوي قيمتُه الواحد، ينتقلُ عبر خط ارتباط مثقّل بوزن أولي وبعدئذ يتمُّ بناءُ الشبكة عبر ربطِ عددٍ من هذه العصبونات ببعضها، وهنا يُمكن تمييزُ طريقتين مختلفتين للقيام بذلك :

- ✓ شبكة التغذية الأمامية (Networks Neural Forward Feed): يتم استقبال كلُّ وحدةٍ المعلوماتِ من الوحداتِ السابقة وتُوصلُها للوحداتِ التالية فلا يُمكن أن تعود المعلوماتُ بالاتجاه المعاكس، وعادةً مايتمُّ ترتيبُ الوحداتِ في طبقات، إذ تستلمُ الوحداتُ المعلومات من الوحدات في الطبقة السابقة ،قد تتكوّن الشبكة من طبقةٍ من الوحدات تصلُ الدخلَ بالخرج وتُدعى بالشّبكة ذاتِ الطبقة الواحدة، أو قد تتضمنُ عدةَ طبقاتٍ من الوحدات وتُسمى بالطبقاتِ الخفيّة
- ✓ شبكة التغذية المتكرّرة(Networks Neural Back Feed): بعد أن تتمَّ معالجةُ المعلومات يتمُّ إعادة إدخالِها من جديدٍ إلى الوحدات لتتمَّ مُعالجتُها مرَّةً أُخرى وهي أشبه بالذاكرة قصيرة الأمد في دماغنا.



الشكل (١) نموذج الشبكة العصبونية

3-2-2 أنواع تعلم الشبكات العصبونية:

- ✓ التعلم الخاضع للإشراف (Supervised Learning): في هذا النوع من التعلم يتم تدريب الشبكة باستخدام دخل معروف الخرج مسبقا الهدف من ذلك هو تعلمُّ كيفية ربط الدخل مع الخرج ليصبح بالإمكان مستقبلًا التنبؤ بالخرج من أجل أي دخل جديد. يندرج تحت هذا النوع أنواعٌ فرعية من التعلم بحسب الخرج المطلوب من أهم هذه الأنواع:

- التصنيف (Classification): وهو النوع الأكثرُ استخدامًا، في هذا النوع يكون الدخل مصنفًا إلى نوعين أو أكثر. وهدفُ عملية التعلم إنتاجُ نموذجٍ يستطيع تصنيفَ أي دخل جديد إلى نوع أو أكثر من الأنواع المعرفة سابقًا.
- الانحدار (Regression): هذا النوع شبيه بالتصنيف، إلا أنه يتنبأ بقيم مستمرةٍ للتصنيف بدلًا من أصنافٍ منفصلة.

✓ التعلم غير الخاضع للإشراف (Unsupervised learning):

بعكس النوع السابق فإن هذا النوع من التعلم يتم تدريب الشبكة عن طريق بياناتِ الدخل بدون أي خرجٍ معرّف مسبقًا، والهدف هنا هو استنباطُ نماذجَ جديدة وعلاقاتٍ خفية بين البيانات[7] .

3-2-3 توابع التفعيل (Activation functions):

توابع التفعيل [7] هي معادلات رياضية تحدد خرج الشبكة العصبونية بحيث يرتبط التابع بكل خلية عصبية في الشبكة، وتحدد ما إذا كان ينبغي تنشيطها أم لا، وذلك بالاعتماد على مدخلات كل خلية عصبية لها صلة بتنبؤ النموذج .تساعد توابع التفعيل أيضًا على جعل قيمة ناتج كل خلية عصبية يتراوح بين [0,1] أو بين [1,1-].

بالإضافة الى أن توابع التفعيل يجب أن تكون فعالة من الناحية الحسابية لأنها تحسب عبر آلاف أو حتى ملايين الخلايا العصبية لكل عينة بيانات وتقسم هذه التوابع الى توابع خطية وتوابع لا خطية.

3-2-4 خوارزميات التعلم:

❖ **Backpropagation:**

تستخدم في تدريب الشبكات العصبونية كاملة الارتباط وذات التغذية الأمامية ومتعددة الطبقات وغير الخطية، وتعتبر هذه الخوارزمية تعميم لطريقة التدريب بنمط تصحيح الخطأ. ويتم تنفيذ هذه الخوارزمية من خلال مرحلتين رئيسيتين هما:

١. مرحلة الانتشار الأماميFeed forward Propagation
٢. مرحلة الانتشار العكسيBack Propagation

✓ مرحلة الانتشار الأماميFeed forward Propagation :

تبقى الأوزان كما هي وتبدأ هذه المرحلة بدخل الشبكة، حيث يتم تخصيص كل عنصر معالجة من طبقة عناصر الإدخال لأحد مكونات الشعاع الذي يمثل الدخل، وتستجيب وحدات طبقة الإدخال لقيم مكونات متجهات الدخل ويعقب ذلك انتشار أمامي عبر بقية طبقات الشبكة.

✓ مرحلة الانتشار العكسيBack Propagation :

وهي مرحلة ضبط أوزان الشبكة، إن خوارزمية الانتشار العكسي القياسية هي خوارزمية الانحدار التدريجي (gradient descent algorithm)والتي تسمح لأوزان الشبكة أن تتحرك على الجانب السلبي من تابع الأداء. إن دور الانتشار العكسي يعود إلى الطريقة التي يتم بها حساب الانحراف (الخطأ) لطبقات الشبكة المتعددة اللاخطية، حيث يتم في أحد مراحل التعليم إعادة انتشار الإشارة من الخرج إلى الدخل بشكل عكسي، ويتم خلالها ضبط أوزان الشبكة[7]

❖ **خوارزمية(LM) Levenberg-Marquardt:**

توفر هذه الخوارزمية حلاً عدديًا لمشكلة الوظيفة غير خطية بشكل سريع وبتقارب مستقر، في مجال الشبكات العصبية الاصطناعية، تعد هذه الخوارزمية مناسبة لتدريب المشكلات الصغيرة والمتوسطة[8].

تدمج خوارزمية Levenberg–Marquardt بين:

- ✓ طريقة steepest descent
- ✓ خوارزميةGauss–Newton

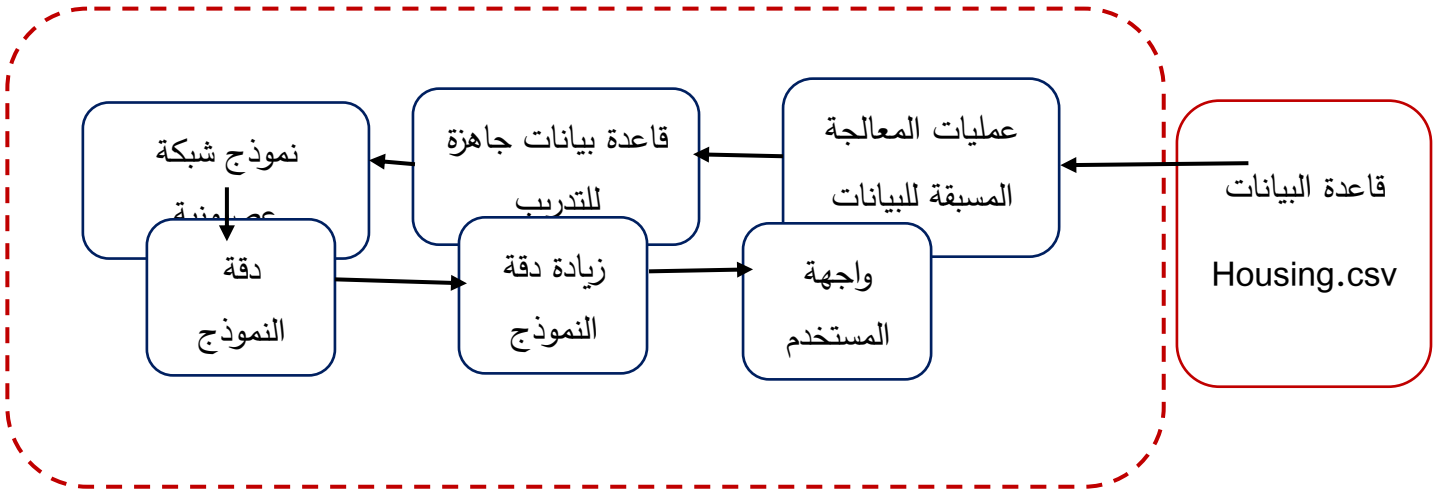
ولكن LM يتميز بالسرعة والاستقرار الموجودة في خوارزمية Gauss–Newton أكثر من طرق steepest descent، إلا أن LM أكثر قوة من خوارزميةGauss–Newton ، وذلك يعود إلى أن الكثير من الحالات يمكن أن تتقارب بشكل جيد حتى لو كان تزايد الخطأ تربيعياً.

على الرغم من أن خوارزمية Levenberg–Marquardt تميل إلى أن تكون أبطأ قليلاً من خوارزمية Gauss–Newton(في الوضع المتقارب)، إلا أنها تتقارب بشكل أسرع بكثير من طريقة steepest descen.

4–المناقشة:

نُفذ البحث بتطبيق سلسلة من الخطوات على قاعدة البيانات لتصبح مهيئة للاستخدام ضمن نموذج الذكاء الاصطناعي ثم بناء شبكة عصبونية اصطناعية للتنبؤ بشكل تلقائي بقيمة العقار ثم العمل على تحسين دقة التنبؤ للنموذج بأعلى دقة ممكنة.

المخطط التالي المبين بالشكل (٢) يصف آلية تنفيذ البحث



الشكل (٢) خطوات البحث

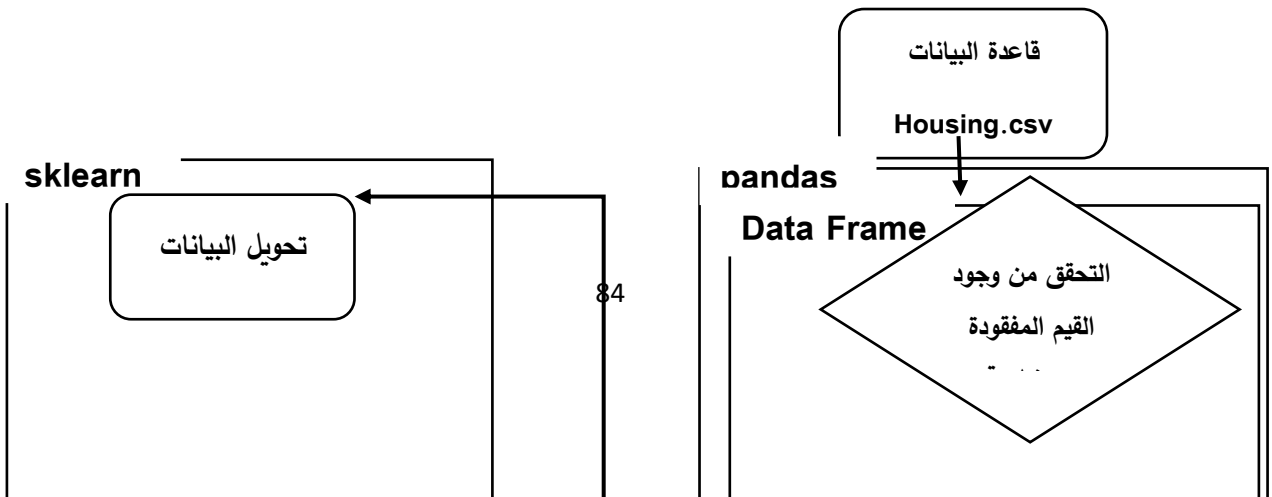
تكمن أهمية البحث من خلال الحصول على قيمة العقار بشكل مؤتمت وبالاعتماد على واجهة تطبيق عبر الانترنت خلال فترة زمنية قصيرة.

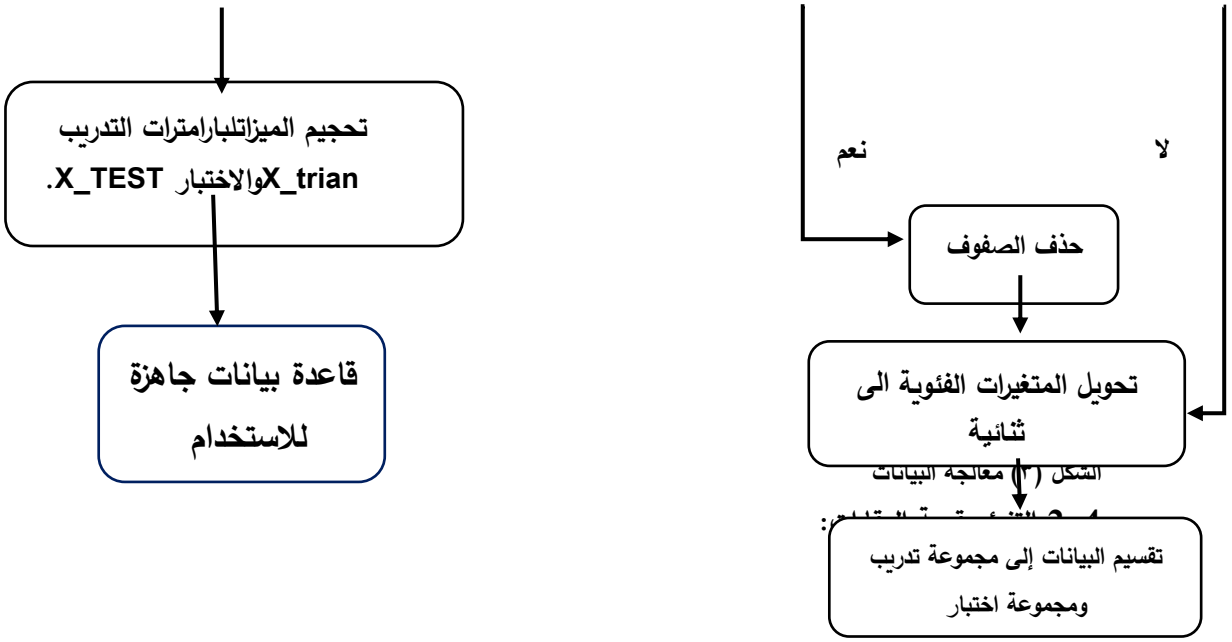
4–1العمليات المعالجة المسبقة للبيانات:

تعتبر عملية معالجة البيانات الخطوة الأساسية والاولية والضرورية في بناء أي نموذج من نماذج الذكاء الاصطناعي وتكمن أهميتها بتأثيرها بشكل مباشر على موثوقية ودقة النموذج المقترح.

ضمن بحثنا تم اتباع مجموعة من الخطوات الشكل (٣) للوصول الى قاعدة بيانات مهيئة للاستخدام والتدريب من قبل الشبكة العصبونية الاصطناعية:

- تضمين مكتبةpandas لقراءة ملف CSV بالاعتماد على تابع read_csv على شكل Data Frame(هيكل بيانات ثنائي الأبعاد يستخدم لتخزين وتنظيم البيانات).
- تنظيف البيانات من خلال:
  - التحقق من وجود القيم المفقودة بالاعتماد على مكتبة pandas من خلال هيكل البيانات Data Frame.
  - حذف الصفوف التي تتضمن قيم مفقودة بالاعتماد على التابع dropna.
  - تحديد الصفوف التي تحوي على قيم متشابهة بالاعتماد على التابع duplicatedوحذفها من خلال التابع drop_duplicates
- تنسيق البيانات بشكل قابل للاستخدام من خلال:
  - تحديد المتغيرات الفئوية (categorical variables)ضمن قاعدة البيانات التي تحتوي القيم "yes" و"no" .
  - التحويل الى المتغيرات الثنائية (binary variables) بالاعتماد على تابع lambd المحتوى ضمن هيكل البيانات Data Frame والذي يعمل على مقارنة قيم العمود مع القيمة "yes"، في حال المساواة يتم تحويلها إلى القيمة ١ وإذا كانت القيمة تساوي"no" ، يتم تحويلها إلى القيمة ٠.
  - تعيين بارامترات التدريب(X) وبارامتر الهدف (Y) وتقسيم البيانات إلى مجموعتين: مجموعة تدريب ومجموعة اختبار بنسبة ٧٠:٣٠.
- تحويل البيانات من خلال:
  - تحجيم الميزات بشكل يصبح متوسط العينة مساوياً لقيمة الصفر والانحراف المعياري مساوياً لقيمة الواحد.
  - تم تطبيق تحجيم الميزات بالاعتماد على مكتبة sklearn على بارامترات التدريب X_trian والاختبار X_TEST.





الشكل (٣) معالجة البيانات

ضمن بحثنا قمنا بتدريب شبكة عصبونية اصطناعية لتصنيف بارامترات العقار والتنبؤ بالقيمة النهائية بشكل مؤتمت وتلقائي.

تحتوي قاعدة البيانات التي استخدمت لتدريب الشبكة العصبونية الاصطناعية على (٥٤٦) قيمة لعقارات معتمدة على مجموعة من البارامترات (الموقع، المساحة، الطريق الرئيسي، .....).

4-2-1 الشبكات العصبونية:

التجربة الأولى

تم استخدام شبكة عصبونية RegressorMulti-Layer Perceptronمتعددة الطبقات تتضمن ثلاث طبقات مخفية تحتوي على (١٠٠،١٠٠،٢٠٠) عصبون تجريبياً مع دخل عبارة عن ٩ بارامترات وخرج واحد (سعر العقار) وفق التالي:

- ✓ تم تقسيم البيانات الى مجموعتين: مجموعة تدريب ٧٠% تتضمن تقريباً (٣٨٢)عينة ومجموعة اختبار ٣٠% تتضمن تقريباً (١٦٣) عينة وقد تم الاعتماد على الطريقة الاستمرارية في تقسيم البيانات بحيث تتضمن هذه الطريقة الافتراضات الأساسية حول مجموعة البيانات، بما في ذلك اختيار مجموعة البيانات بشكل عشوائي وتوزيعا بشكل مستقل ومماثل بالإضافة الى أن التوزيع المستقل لا يتغير بمرور الوقت.
- ✓ يستخدم النموذج خوارزمية (Stochastic Gradient Descent)للتدريب على البيانات ويتم تحديث الأوزان والانحراف في النموذج بناءً على الخطأ الذي تم حسابه باستخدام خوارزمية (Backpropagation) بحيث يتم تكرار هذه العملية على البيانات التدريبية لضبط النموذج وتحسين أدائه مع خطوة تعلم (٠,٠١).
- ✓ تابع تنشيط ReLU.

تطلب تدريب الشبكة العصبونية الى ٣٠٠ دورة واعطت النتائج وفق:

- **مصفوفة الارتباك**

تعد مقياس الأداء لمشكلة التنبؤ بحيث تعتبر:

- ✓ التنبؤ الصحيح هو عبارة عن الإيجابيات الحقيقيةTPأو السلبيات الحقيقيةTN.
- ✓ التنبؤ الخاطئ هو عبارة عن الإيجابيات الكاذبةFPأو السلبيات الخاطئةFN.
- ✓ تحسب الدقة [9] وفق العلاقة:$\text{Accuracy}=\frac{TP+TN}{TP+FN+TN+FP} * 100$

- **مخطط Roc**

- ✓ يظهردقة النموذج
- ✓ بحيث يقوم بتطبيق قيم العتبة [0,1] على الخرج.
- ✓ يعطى مخطط Roc وفق معدل الإيجابي الحقيقي (احتمال التنبؤ الصحيح) مقابل معدل الإيجابي الخاطئ (احتمال التنبؤ الخاطئ).

- **Mean Squared Error (MSE)**

- ✓ مقياس يستخدم لقياس مدى الفرق بين القيم الفعلية والقيم المتوقعة في نموذج أو تنبؤات معينة.
- ✓ يحسب مربع الخطأ [10] وفق العلاقة: MSE = (1 − Model Accuracy) / Model Accuracy

جدول (1) نتائج التجربة الأولى

ANN	Accuracy	Roc	MSE
learning_rate_init=0.01 activation=relu	٠,٦٣	٠,٧٨	٠,٥

<u>التجربة الثانية (تحسين دقة النموذج)</u>

تم استخدام شبكة عصبونية RegressorMulti−Layer Perceptronمتعددة الطبقات تتضمن طبقتين مخفيتين تحتوي على (٢٠٠،١٠٠) عصبون تجريبياً مع دخل عبارة عن ٩ بارامترات وخرج واحد (سعر العقار) مع تعديل التالي:

- ✓ تابع تنشيط identity.
- ✓ خطوة تعلم (٠,٠٣)
- ✓

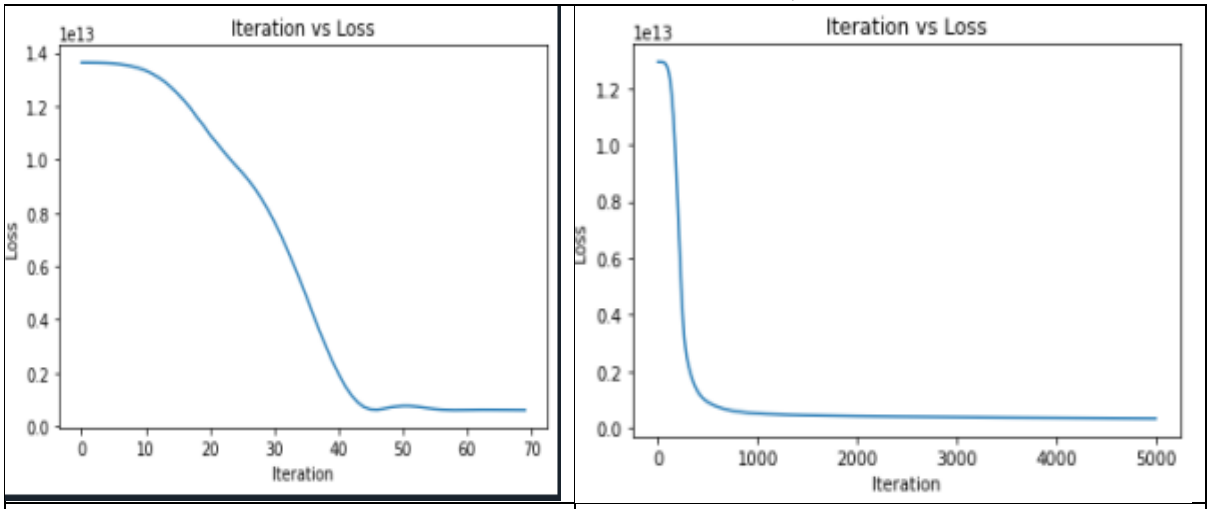
جدول (2) نتائج التجربة الثانية

ANN	Accuracy	Roc	MSE
learning_rate_init=0.03 .activation=identity	0.88	0.88	0.1

4−3 مناقشة النتائج:

يظهر الشكل (٤) مخطط الأداء للشبكة العصبونية في التجريبتين يمكن الملاحظة أن:

- تخفيض عدد الطبقات المخفية مع استخدام تابع التنشيط **identity** (تمرر قيم الدخل الى الخرج دون تعديل)مع خطوة تعلم ٠,٠٣أدى الى تعلم الشبكة عند التكرار ٤٥ مع أفضل قيمة لأداء الشبكة هي ٠,١وهي قريبة من 0مما يدل على دقة تصنيف الشبكة في التجربة الثانية.
- تطلبت الشبكة العصبونية باستخدام ثلاث طبقات دخل وتابع التنشيط relu الى 500 تكرار للتعلم مع أفضل قيمة لأداء الشبكة وهي٠,٥ وذلك ضمن التجربة الأولى.
- تم إعادة التجربة الأولى باستخدام تابع التنشيط relu وطبقتين مخفيتين وخطوة تعلم ٠,٠٣ تطلب تعليم الشبكة ٤٠٠ تكرار مع انخفاض في الدقة الى ٥٩%.
- تم إعادة التجربة الثانية باستخدام تابع التنشيط **identity**وثلاث طبقات مخفية وخطوة تعلم ٠,٠١ تطلب تعليم الشبكة ١٠٠ تكرار مع انخفاض في الدقة ٨٢%.

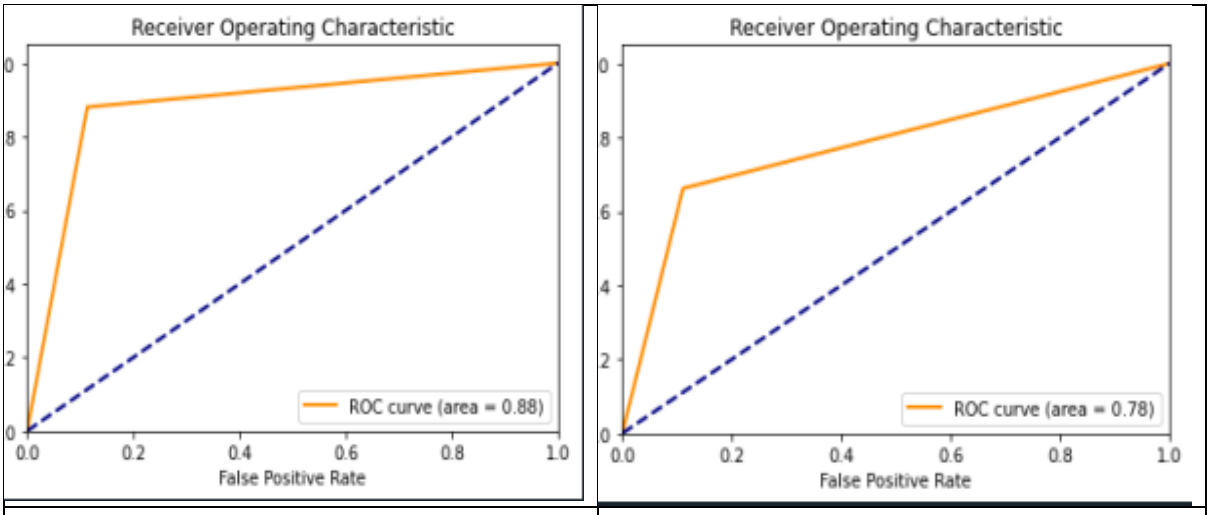


learning_rate_init=0.03/ activation= identity	learning_rate_init=0.01/ activation=relu

الشكل (4) مخطط الأداء

يظهر الشكل (5) مخططٌ ROC في التجريبتين يمكن الملاحظة أن:

- تدريب الشبكة العصبونية في التجربة الثانية أدى الى منحني يقترب من الزاوية اليسرى العلويةبنسبة (٠,١٥)للحالات الصحيحة مقابل الخاطئة مع مساحة تحت المنحنى بقيمة ٠,٨٨تقترب من ١ (كلما زادت المساحة الموجودة تحت المنحنى، كلما كان الاختبار أكثر دقة).
- تدريب الشبكة العصبونية في التجربة الاولى أدى الى منحني يقترب من الزاوية اليسرى العلويةبنسبة (٠,٢٢) للحالات الصحيحة مقابل الخاطئة مع مساحة تحت المنحنى بقيمة ٠,٧٨
- تم الاعتماد على المساحة تحت المنحني كمقياس لأداء النموذج وعليه يعتبر النموذج في التجربة الثانية أكثر دقة للتنبؤ.



learning_rate_init=0.03/ activation= identity	learning_rate_init=0.01/ activation=relu

الشكل (٥) مخطط ROC

يظهر الشكل (٦) مصفوفة الارتباك في التجربتين يمكن الملاحظة أن:

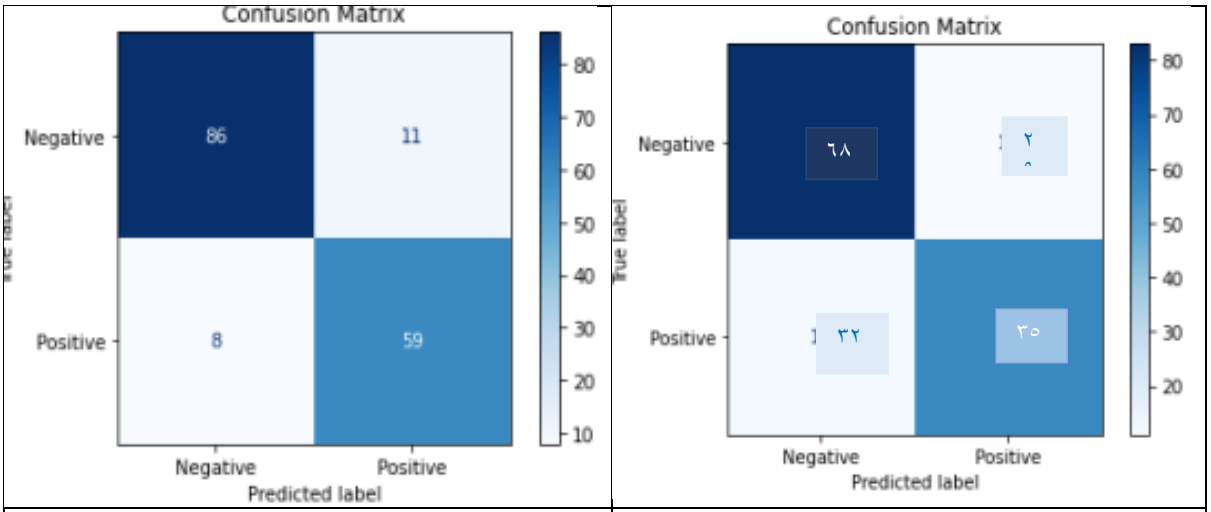
- التنبؤات الصحيحة في التجربة الأولى هي عبارة عن ١٠٤عينة فيما التنبؤات الخاطئة ٦١ عينة وبحساب دقة نموذج:

$$\text{Accuracy}=\frac{TP+TN}{TP+FN+TN+FP}*100=\frac{69+35}{69+29+32+35}*100=63\%$$

- التنبؤات الصحيحة في التجربة الثانية هي عبارة عن ١٤٥ عينة فيما التنبؤات الخاطئة ١٩ عينة وبحساب دقة نموذج:

$$\text{Accuracy}=\frac{TP+TN}{TP+FN+TN+FP}*100=\frac{86+59}{86+11+8+59}*100=88.4\%$$

- النموذج في التجربة الثانية أكثر دقة من التجربة الأولى مع الأخذ بعين الاعتبار أن حساب الدقة تم بالاعتماد على مصفوفة الارتباك لمجموعة الاختبار (١٦٤).

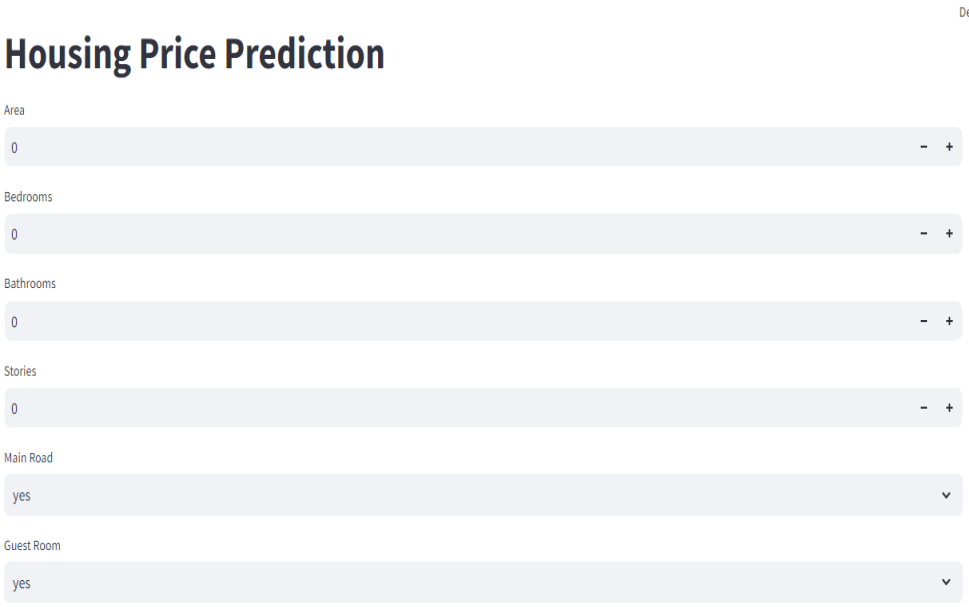


learning_rate_init=0.03/ activation= identity	learning_rate_init=0.01/ activation=relu

الشكل (٦) مصفوفة الارتباك

4-3 الواجهة الرسومية:

تم تصميم واجهة رسومية GUI تفاعلية الشكل (٧) بحيث يمكن للعميل ادخالبارامترات الدخل(الموقع،المساحة، الطريق الرئيسي، .....) والحصول على بارامتر الخرج (قيمة العقار) عبر المتصفح دون الحاجة إلى تثبيت أي برامج أو تطبيقات إضافيةمع الاخذ بعين الاعتبار الاعتماد على مكتبة streamlit في التصميم.



الشكل (٧) التنبؤ بقيمة العقار

٥- الاستنتاجات والتوصيات:

- تطبيق عمليات المعالجة المسبقة على قاعدة البياناتكان له نتيجة فعالة في زيادة كفاءة ودقة النموذج المقترح.
- تنفيذ النموذج بالاعتماد على الشبكة العصبونية الاصطناعية مع دالة التنشيط identity مع خطوة تعلم ٠,٠٣ أدى الى دقة أفضل للتنبؤ بقيمة العقار.
- تنفيذ النموذج المقترح تطلب فترة زمنية قصيرة بالإضافة الى توفير الموارد الحاسوبية المطلوبة (Core i3/MEMORY 4GB).
- استخدام النموذج المقترح ضمن الأصول العقارية يؤدي الى توفير معلومات كاملة قدر الإمكان عن قيمة العقار.

يمكن المتابعة في البحث من خلال:

- تطوير النموذج وتحسين أدائه من خلال العمل على زيادة دقة النموذج بالاعتماد على الشبكات العصبونية العميقة والخوارزمية الجينية.

- استخدام خوارزمية النمل بعد اعتماد دقة معينة لاختيار أفضل مسار لكل عقار ضمن قاعدة البيانات مما يرفع من سوية النموذج.

المراجع

[1] Yadav, A. (2022). Predicting The Housing Price using Artificial Intelligence/ Machine Learning. International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), 2(6), 152-168. DOI: DOI: 10.48175/568. -168. DOI: 10.48175/568.

[٢] Foryś, Iwona. "Machine learning in house price analysis: regression models versus neural networks." Procedia Computer Science 207 (2022): 435-445.

[3] Çılgın, C., & Gökçen, H. (2023). Machine learning methods for prediction real estate sales prices in Turkey. Revista de la construcción, 22(1), 163-177

[4] Yasser, H. Housing Prices Dataset. Kaggle. Published (50:30 PM). Available online: https://www.kaggle.com/datasets/yasserh/housing-prices-dataset?resource=download (Accessed: 10-2-2024).

[5] Gitconnected,. (2022). "Importance of Data Preprocessing in Machine Learning". [Online]. Available: https://levelup.gitconnected.com/importance-of-data-preprocessing-in-machine-learning-17045bc18d01. [Accessed: 21-2-2024].

[6] Finlay, J. (2020). An introduction to artificial intelligence. Crc Press.

[7] Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. Journal of microbiological methods, 43(1), 3-31.

[8] Smith, J.S., Wu, B. and Wilamowski, B.M., 2018. Neural network training with Levenberg–Marquardt and adaptable weight compression. IEEE transactions on neural networks and learning systems, 30(2), pp.580-587.

[9] Li, H., Giger, M.L., Huo, Z., Olopade, O.I., Lan, L., Weber, B.L. and Bonta, I., 2004. Computerized analysis of mammographic parenchymal patterns for assessing breast cancer risk: effect of ROI size and location. Medical Physics, 31(3), pp.549-555

[10] Charkhi, A., Claeskens, G., & Hansen, B. E. (2016). Minimum mean squared error model averaging in likelihood models. Statistica Sinica, 809-840.