

التنبؤ بالعوائد السكنية لدعم اتخاذ القرار باستخدام خوارزميات شجرة القرار (دراسة حالة محلية أمبدة)

(Predicting Housing Returns to Support Decision-Making Using
Decision Tree Algorithms: A Case Study of Umbadda Locality)

إعداد:

- د. سيف الدين فتوح عثمان (استاذ علوم الحاسوب المشارك بكلية الامارات للعلوم والتكنولوجيا)

profsief@wnu.edu.sd

- د. عبدالباسط محمد حامد (استاذ مساعد علوم الحاسوب بجامعة النيل الابيض)

- أ. محمد ابراهيم محمد الحسن ابراهيم

المستخلص

تهدف هذه الدراسة إلى تحليل العوائد السكنية في منطقة محلية أمبدة باستخدام تقنيات تنقيب البيانات، مع التركيز على تطبيق خوارزميات شجرة القرار لاستخلاص الأنماط والعوامل المؤثرة في قيمة العوائد. تم استخدام برنامج (WEKA) لتطبيق خوارزمية (Random Tree)، وتم تحليل أداء هذه الخوارزمية. أظهرت النتائج أن خوارزمية (Random Tree) حققت أفضل أداء من حيث الدقة والقدرة التنبؤية. توصلت الدراسة إلى أن العوامل المؤثرة في العوائد السكنية تشمل توفر الماء والكهرباء وبقية الخدمات. توصي الدراسة باستخدام نماذج شجرة القرار كأداة داعمة في عمليات اتخاذ القرار الاستثماري والتخطيط العمراني في المنطقة، لما توفره من قدرة على تفسير العلاقات بين المتغيرات والتنبؤ بالعوائد بدقة عالية.

الكلمات المفتاحية: شجرة القرار ، (WEKA)، التنبؤ ، تحليل البيانات، اتخاذ القرار.

Abstract

This study aims to analyze residential returns in the Umbadda locality using data mining techniques, with a focus on applying decision tree algorithms to extract patterns and identify the factors influencing return values. The (WEKA) software was used to apply the (Random Tree) algorithm, and its performance was analyzed. The results showed that the (Random Tree) algorithm achieved the best performance in terms of accuracy and predictive capability. The study found that the key factors affecting residential returns include the availability of water, electricity, and other services. The study recommends the use of decision tree models as a supportive tool in investment decision-making and urban planning in the area, due to their ability to interpret relationships between variables and accurately predict returns.

Keywords: Decision Tree, WEKA, Prediction, Data Analysis, Decision Making.

المقدمة

تعتبر العوائد السكنية من المؤشرات الرئيسية في تقييم الاستثمارات العقارية، حيث تعتمد على عوامل متعددة مثل الموقع، البنية التحتية، والعرض والطلب. في هذه الورقة، يتم تطبيق نموذج شجرة القرار لتحليل العوائد السكنية في منطقة محلية أمبدة، بهدف تقديم رؤى دقيقة تساعد أصحاب القرار في تحسين قراراتهم. يعتمد التحليل على بيانات واقعية، مع استخدام تقنيات متقدمة لتنقيب البيانات لتحسين دقة التنبؤ.

مشكلة الدراسة: تتلخص المشكلة في عدم وجود آلية علمية موثوقة تطور وتحسن من اتخاذ القرار في منح تصاديق البناء وايضا وضع الموازنات الايرادية الخاصة بالعوائد السكنية المستقبلية وامكانية التنبؤ بهذه العوائد.

هدف الدراسة: تهدف هذا الدراسة إلى تطوير وتحسين عملية اتخاذ القرار في محلية أمبدة من خلال التنبؤ بإيرادات العوائد السكنية وتحديد العوامل المؤثرة فيها، وذلك باستخدام تقنيات تنقيب البيانات.

الفجوة البحثية

هنالك العديد من الفجوات البحثية في هذا المجال ، تهدف هذه الدراسة الي سدها ، مثل :

- العوامل التي تؤثر علي زيادة العوائد السكنية بالمحلية .
- الطريقة العلمية في منح تصاديق البناء ونوعها بالمحلية.
- الطريقة العلمية الموثوقة لتوزيع الخدمات بالمحلية.

منهجية البحث

تم اتباع المنهج الوصفي التحليلي في مساحة مقدره من هذه الدراسة والمنهج شبه التجريبي لما تقتضيه هذه الدراسة من إجراء التجربة العملية بمحلية امبدة ، حيث يتكون من المراحل التالية:

اولا : تحديد المشكلة وأهداف البحث

- تحديد العوامل المؤثرة على رفع قيمة العوائد السكنية.
- استخدام تقنيات تنقيب البيانات لتطوير نموذج تنبؤي يساعد في اتخاذ القرار.

ثانيا : جمع البيانات

- الحصول على بيانات العوائد السكنية من الجهات المختصة بمحلية أمبدة.

ثالثا : معالجة البيانات

- تنظيف البيانات من القيم المفقودة أو المتطرفة لضمان دقة التحليل.

رابعاً : اختيار وتطبيق تقنيات تنقيب البيانات

- استخدام خوارزمية التنبؤ (Random Tree) .
- تقييم أداء النماذج باستخدام مقاييس مثل متوسط الخطأ المطلق (MAE).

خامساً : تحليل النتائج وتفسيرها

سادساً : تقديم التوصيات واتخاذ القرار

الدراسات السابقة:

هنالك العديد من الدراسات ذات صلة مباشرة او غير مباشرة بموضوع الدراسة والتي أمكن الحصول عليها اتفقت معنا في المنهجية والموضوع واختلفت بعضها في ادوات التنفيذ ، حيث تناولت الدراسات مواضيع الذكاء الاصطناعي و تنقيب البيانات ودورها في تطوير وتحسين اتخاذ القرار. نذكر منها :

الدراسة الاولى : (محمد سالم ، 2023) ملخص الدراسة: استهدفت هذه الدراسة تقييم تأثير سياسات الحكومة على سوق العقارات السكنية. تم تحليل السياسات الحكومية وتأثيرها على العرض والطلب في السوق باستخدام ادوات تنقيب البيانات.

النتائج: السياسات الحكومية مثل تخفيض الضرائب وزيادة الحوافز أدت إلى زيادة في بناء العقارات السكنية. التوصيات: استمرار دعم السياسات التي تشجع على الاستثمار في القطاع العقاري السكني.

الدراسة الثانية: (بشري عيد ، 2023) ملخص الدراسة: تناولت هذه الدراسة دور الذكاء الاصطناعي في التنبؤ بالأسواق العقارية، مع مقارنة بين الأساليب التقليدية وأساليب الذكاء الاصطناعي. تم التركيز على تقنيات مثل التعلم الآلي وتحليل البيانات الكبيرة.

النتائج: الذكاء الاصطناعي يمكنه معالجة كميات ضخمة من البيانات وتقديم تنبؤات دقيقة، يتميز الذكاء الاصطناعي بالقدرة على التكيف المستمر والتنبؤ بالآزمات والتغيرات المفاجئة. التوصيات: الاستثمار في تقنيات الذكاء الاصطناعي لتحسين دقة التنبؤات العقارية، تدريب فرق العمل المتخصصة على استخدام هذه التقنيات بفعالية.

الدراسة الثالثة : (علي محمد ، 2021) ملخص الدراسة: هدفت هذه الدراسة إلى تحليل تأثير جائحة كوفيد-19 على أسعار العقارات في المناطق الحضرية. تم جمع بيانات من عدة مدن لتحليل التغيرات في الأسعار. النتائج: انخفاض أسعار العقارات في المناطق التجارية وزيادة الطلب على العقارات السكنية في الضواحي. التوصيات: إعادة تقييم استراتيجيات الاستثمار العقاري لتتناسب مع التغيرات في الطلب.

الخلفية النظرية والادوات المستخدمة :

تنقيب البيانات (Data Mining) (السيد، أ. (2020)) :

تنقيب البيانات هو العملية التي تهدف إلى اكتشاف الأنماط والمعرفة من خلال تحليل كميات ضخمة من البيانات غير المنظمة أو المنظمة باستخدام تقنيات الذكاء الاصطناعي، والتعلم الآلي، والإحصاء.

مفهوم خوارزمية شجرة القرار (Random Tree) (بكة، م. 2023):

خوارزمية (Random Tree) هي نوع من خوارزميات شجرة القرار تعتمد على إنشاء شجرة باستخدام اختيار عشوائي للخصائص عند كل عقدة. تختلف عن شجرة القرار التقليدية في أنها لا تستخدم جميع الخصائص لتقسيم البيانات، بل تختار مجموعة فرعية عشوائية منها. تُستخدم هذه الطريقة لتحسين التنوع وتقليل التحيز، تمتاز بسرعتها وبساطتها وفعاليتها في التعامل مع البيانات الكبيرة والمعقدة.

الاساس الرياضي لخوارزمية (Random Tree) (Han, J., Kamber, M., & Pei, J. (2012)).

تعتمد خوارزمية (Random Tree) على مبدأ شجرة القرار التقليدية، لكنها تضيف عنصر العشوائية لزيادة التنوع وتحسين الأداء،

في كل عقدة من الشجرة:

- لا يتم النظر إلى جميع السمات الخصائص F ، بل يتم اختيار مجموعة فرعية عشوائية $f \subset F$
- يتم اختيار السمة الأفضل من هذه المجموعة الفرعية باستخدام معيار إحصائي مثل كسب المعلومات أو انخفاض التباين.

أولاً: اختيار السمة باستخدام كسب المعلومات: (Information Gain)

$$\bullet \quad IG(S, A) = Entropy(S) - \sum (|S_v| / |S|) * Entropy(S_v)$$

حيث :

- العقدة عند البيانات مجموعة S .
- للتقسيم المرشحة : السمة A .
- A للسمة v القيمة تملك التي البيانات : مجموعة S_v .

ثانيا الإنتروبي (Entropy) :

- $Entropy(S) = - \sum (p_i * \log_2(p_i))$

حيث :

- p_i : البيانات مجموعة داخل فئة كل احتمال.

ثالثا في حالة التنبؤ العددي (Regression) :

يُستخدم متوسط مربعات الخطأ (MSE) كالآتي:

- $MSE = (1/n) * \sum (y_i - \hat{y}_i)^2$

حيث :

- y_i : الحقيقية القيمة.
- $\hat{y}_i$: المتوقعة : القيمة.
- n : عدد عناصر في البيانات عناصر : عدد

تتميز (Random Tree) بقدرتها على التعامل مع البيانات الكبيرة والمتعددة الأبعاد، كما تُعتبر فعالة في تقليل التحيز وزيادة التباين، مما يجعلها مناسبة للتطبيقات التنبؤية المختلفة.

سبب استخدام خوارزمية (Random Tree) في هذه الدراسة

تم اعتماد خوارزمية Random Tree نظراً لقدرتها على التعامل بكفاءة مع البيانات متعددة الأبعاد، وسهولة تفسير نتائجها، إضافةً إلى دقتها العالية في التنبؤ بفضل استخدامها للخصائص العشوائية في اختيار السمات، مما يُساهم في تقليل التحيز وتحسين الأداء العام للنموذج حيث أنها تعتبر من الخوارزميات المحسنة.

التطوير الإداري (بيتر دراكر، (2021))

التطوير الإداري هو عملية مستمرة تهدف إلى تحسين الأداء التنظيمي وزيادة كفاءة المؤسسات. تعتمد فعاليته على التزام القيادة، تبني التكنولوجيا، وتطوير الموارد البشرية. ومع استمرار التحديات، فإن الابتكار والتخطيط الاستراتيجي يمثلان مفتاح النجاح في تنفيذ التطوير الإداري بكفاءة.

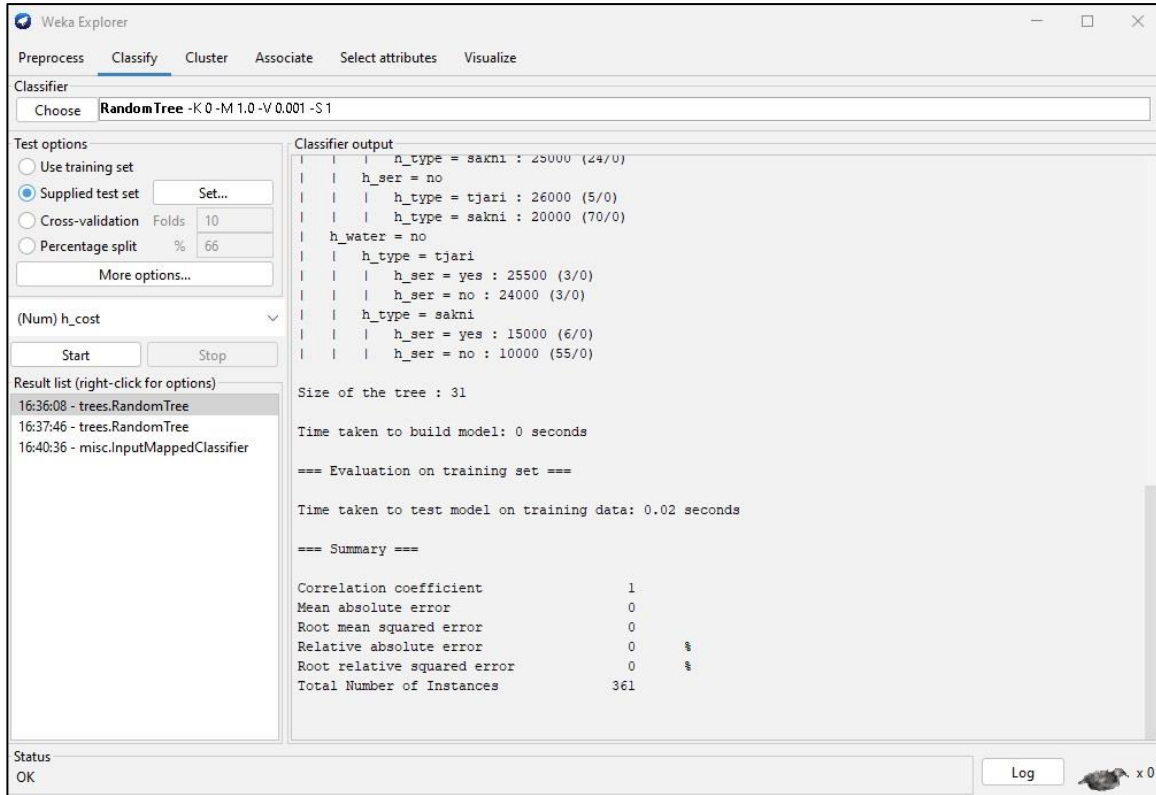
مفهوم اتخاذ القرار وأهميته (يوسف، ي، 2019)

اتخاذ القرار هو عملية اختيار البديل الأفضل من بين عدة بدائل متاحة، بناءً على تحليل البيانات، التنبؤ بالنتائج، وتقييم المخاطر. يعتبر اتخاذ القرار عنصرًا أساسيًا في نجاح أي مؤسسة، حيث يؤثر بشكل مباشر على كفاءة الأداء، جودة الخدمات، وتحقيق الأهداف الاستراتيجية.

برنامج ويكا للتنقيب في البيانات (فتيحة بومرين، 2015)

تعتبر (WEKA) واحدة من أكثر الأدوات شيوعًا في مجال تعلم الآلة وتنقيب البيانات بفضل سهولتها وفعاليتها. تتيح هذه الأداة للمستخدمين تجربة خوارزميات مختلفة، وتقديم تقارير مفصلة عن الأداء، وفحص البيانات وتجهيزها للمعالجة.

تدريب وتقييم الخوارزمية علي عينة البيانات التاريخية بالمحلية ومناقشة النتائج



شكل رقم 1: شاشة عرض ملخص عمل خوارزمية (Random tree)

- بعد تشغيل الخوارزمية علي البيانات الاصلية ظهر لنا في النتائج ان معامل الارتباط يساوي الواحد الصحيح وهو المعامل الذي يقيس مدي قوة العلاقة بين القيم الحقيقية والقيم المتوقعة، حيث تتراوح

قيمتها عادة بين -1 و +1 ، وكلما اقتربت قيمتها من الواحد كان النموذج دقيق جدا ، وإذا كانت مساوية للصفر فذلك يعني ان النموذج عشوائي ولا يتنبأ بشكل جيد، وإذا كانت قريبه من -1 فان النموذج يعكس العلاقة بشكل خاطئ، وعليه فان النموذج هنا لديه دقة جيدة جدا في التنبؤ بالقيم العددية.

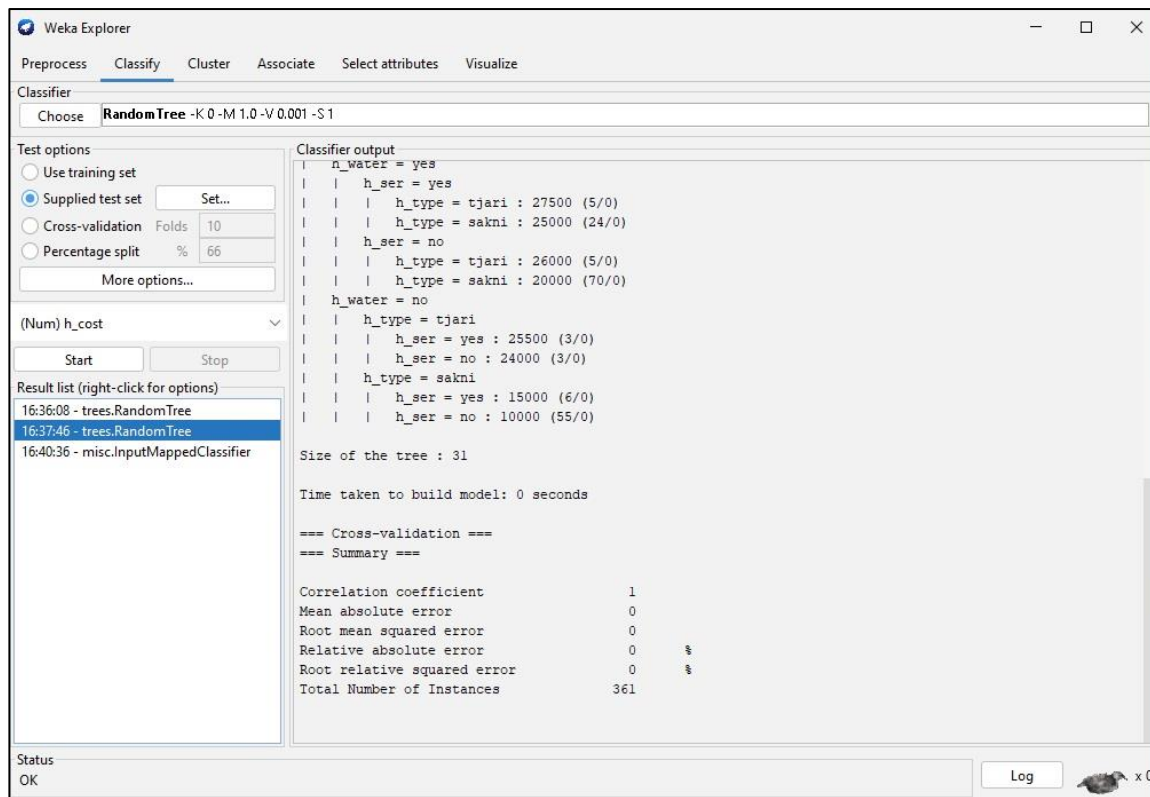
- وإذا نظرنا الى الخطأ المطلق المتوسط (Mean Absolute Error - MAE) والذي هو عبارة عن متوسط الفرق المطلق بين القيم الحقيقية والقيم التي توقعها النموذج الذي بدوره كلما كان صغيرا كان النموذج افضل ، وجدناه هنا مساويا للصفر وذلك يعني ان متوسط خطأ التوقعات هو صفر عن القيم الحقيقية وهذه تعتبر افضل نتيجة .

- وإذا نظرنا الى الجذر التربيعي لمتوسط مربعات الخطأ (Root Mean Squared Error - RMSE) والذي بدوره يقيس مدى تباعد القيم المتوقعة عن القيم الحقيقية ، مع اعطاء اهمية اكبر للأخطاء الكبيرة ، وعندما يكون ال (RMSE) مرتفعا فقد يكون هنالك خطأ كبير في بعض التوقعات الفردية، وهنا نجده ايضا مساويا للصفر مما يدل ان انحراف القيم المتوقعة عن الحقيقية يكاد يكون منعدم ، وهي من افضل النتائج .

- وإذا نظرنا الى نسبة الخطأ المطلق النسبي (Relative Absolute Error - RAE) والتي تقارن بين ال (MAE) وخطأ نموذج بسيط يعتمد فقط على المتوسط العام للبيانات ، وإذا كان ال (RAE) قريبا من صفر% فالنموذج افضل من التوقعات العشوائية ، وعندنا هنا يعتبر افضل لانه مساويا لل صفر%

- وإذا نظرنا الى نسبة الجذر التربيعي لمتوسط مربعات الخطأ النسبي (- Root Relative Squared Error - RRSE) ، وهو مشابه ايضا لل (RMSE) لكنه يحسب كنسخة نسبية مقارنة بنموذج بسيط يعتمد على المتوسط ، وكلما قل كان النموذج افضل ، وهنا جاء مساويا للصفر مما يدل على ان النموذج دقيق لحد كبير .

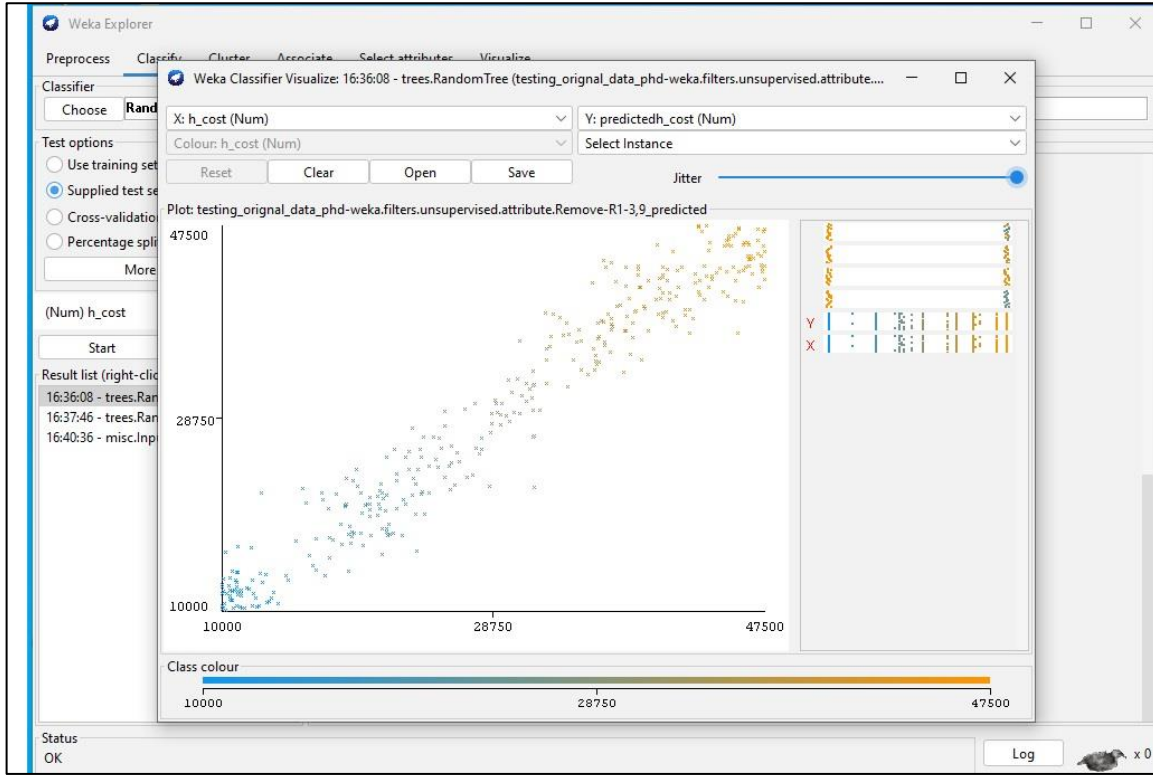
- اخيرا تظهر عدد العينات المستخدمة (Total Number Of Instances) ، والتي كلما كانت كبيرة زادت دقة النموذج بشرط ان تكون البيانات متنوعة وكافية .



شكل رقم 2: شاشة عرض ملخص عمل تقييم الخوارزمية

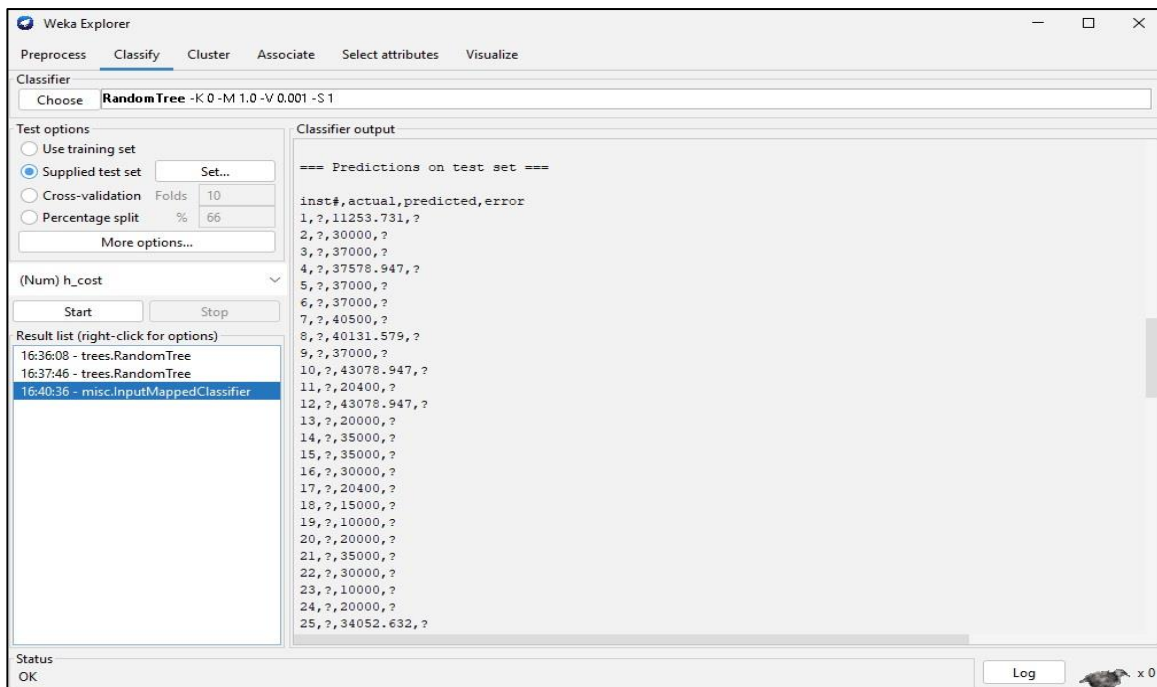
- في هذه الشاشة تم استخدام احد خيارات تقييم الخوارزمية وهو (Cross-Validation) الذي يقوم بتقسيم البيانات الي تدريب واختبار ، وعند ملاحظة القراءات في قسم الملخص بعد تشغيله علي البيانات نجد انه جاء مشابهة تماما ملخص الخوارزمية في الشكل السابق (5-3-1) وهذا يدل ان عملية التقييم اثبتت فعالية الخوارزمية تماما ، وانها صالحة وجاهزة لاي عملية استنتاج او تنبوء .

التنبؤ بالعوائد السكنية المستقبلية لعينة اختبار ومناقشة النتائج



شكل رقم 4 : شاشة معاينة الاخطاء لخوارزمية (Random Tree)

- يتم عبرها معاينة عملية الاخطاء بعد تطبيق الخوارزمية ، وملاحظة القيم الشاذة ، وايضا عبرها يتم حفظ بيانات التنبؤ الجديدة في ملف خارجي.
- واذا نظرنا الي الشكل جيدا فاننا نجد انه لا توجد اي بيانات متطرفة علي الرسم او اي قيم شاذة . وكل القيم تتراوح بين فئات العوائد السكنية ، وذلك ان دل فانما يدل علي شيئين :
الاول : ان البيانات تم تجهيزها جيدا من حيث القيم المتطرفة او المفقودة
الثاني : ان الخوارزمية قد ادت دورها بكل دقة في عملية التحليل والتنبؤ



شكل رقم 5 : شاشة اجراء التنبؤ لخوارزمية (Random Tree)



شكل رقم 6 : ملف مخرجات التنبؤ لخوارزمية (Random Tree)

- نلاحظ في الشكلين اعلاه بعد تنفيذ الخوارزمية واختبارها قمنا بتشغيلها علي عينة اختبار خارجية تحتوي علي مائة منزل غير محدد العوائد السكنية لهم ، علي اساس ان تقوم الخوارزمية بعملية التنبؤ بالقيم

- المطلوبه ، وقد قامت باستنتاج كل القيم الجديدة داخل البرنامج في الشكل (1) لنقوم بحفظها بملف خارج البرنامج كما بالشكل (2) ، وعند تحليل هذه القيم الجديدة ومقارنتها مع القيم الحقيقية السابقة وجدنا ان هذه الاستنتاجات ممتازة جدا ولايوجد بها اي انحراف كبير يذكر وانت بنتيجة ممتازة جدا .
- عند تحليل النتائج بالشكلين اعلاه وجدنا ان هنالك ارتباط كبير ومؤثر جدا في قيمة العوائد السكنية تقوم به بيانات توفر الخدمات مثل الماء والكهرباء والمستشفيات وبقية المرافق ، حيث انه كلما كان المنزل قريب منهم فان عوائد السكنية تكون عالية ، وهذا تناسب طردي.

الخاتمة والنتائج

- الحصول علي نتائج التنبؤ بقيمة العوائد السكنية المستقبلية مما يساعد متخذي القرار في اتخاذ القرارات المالية والموازنات الايرادية بالمحلية.
- توفر الماء والكهرباء والخدمات المتوفرة كانا من العوامل الأكثر تأثيراً على العوائد السكنية.
- عدم الجدوي من منح تصاديق البناء في المناطق التي ليس بها خدمات، لان التنبؤ بالعوائد السكنية فيها اظهر ضعف واضح في الايرادات.
- كل ذلك يساعد متخذي القرار ويقوم بتطوير قراراتهم علي اسس علمية موثوقة.

التوصيات

- دمج بيانات زمنية لتقييم تغير العوائد السكنية بمرور الوقت.
- تجربة نماذج هجينة تجمع بين شجرة القرار والشبكات العصبية.
- تطبيق النموذج على مناطق أخرى للتحقق من قابلية التعميم.

المراجع

1. السيد، أ. (2020). 'تحليل وتنقيب البيانات في نظم المعلومات'. الطبعة الثانية. القاهرة: دار الفكر.
2. فتيحة بوهرين، ف. (2015)، 'ورقة عمل عن التنقيب في البيانات باستخدام ويكا' ، جامعة قسنطينة.الجزائر.
3. بكة، م. (2023) 'تنقيب البيانات: المراحل والأدوات والخوارزميات والكتب'.
4. بيتر دراكر، ب. (2021) ، 'التطوير الإداري الحديث'.
5. يوسف، ي. (2018). 'مقدمة في نظم المعلومات الإدارية'. الرياض: دار الخريجي.
6. Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts