بناء شجرة قرارات تصنيفية كأداة تنقيب عن البيانات Data Mining للتنبؤ بمستوى الاحتفاظ بالأقساط في سوق التأمين المصري باستخدام خوارزمية C4.5

د. أسامة ربيع أمين سليمان *

* **د. أسامة ربيع أمين سليمان** أستاذ التأمين المساعد بقسم الإحصاء والرياضيات والتأمين كلية التجارة – جامعة مدينة السادات.

Email:

الملخص:

يتمثل الهدف الأساسي من هذه الدراسة في تحليل العوامل والمتغيرات المؤثرة على مستوى الاحتفاظ بالأقساط في سوق التامين المصري، من أجل بناء شجرة قرارات تصنيفية مبنية على خوارزمية C4.5 للتنبؤ بمستوى الاحتفاظ في سوق تأمينات الممتلكات والمسؤولية في سوق التأمين المصري، كأحد أدوات التنقيب عن البيانات التي تتميز بسهولة الفهم والتطبيق، وعدم الحاجة إلى شروط نظرية معقدة كما هو الحال في النماذج المعلمية وخاصة النماذج الخطية العامة. وقد توصلت الدرسة إلى أن أهم المتغيرات التي يمكن الاعتماد عليها في التنبؤ بمستوى الاحتفاظ بالأقساط: عدد شركات التأمين و الرافعة المالية لسوق التأمين، حيث أن زيادة عدد شركات التأمين في سوق التأمين المصري من شأنه أن يساعد على زيادة الاحتفاظ بالأقساط. كما أن انخفاض الرافعة المالية ودرجة مخاطر التمويل تساعد أيضا على زيادة مستوى الاحتفاظ بالأقساط في سوق التأمين المصري.

Abstract:

The main objective of this study is to analyze determinants affecting the level of retention of premiums in the Egyptian insurance market in order to design a classification decision tree, to predict retention level of premiums in the Egyptian property and liability Insurance Market, as a data mining tool that is characterized with simplicity, easy to apply and required less conditions as in parametric models (GLM). The main conclusion of this paper that increasing the number of insurers and low insurance leverage may increase the retention in the Egyptian insurance market.

١. مقدمة:

ارتبط عدد كبير من التغيرات التي حدثت في النظام التشريعي عام ١٩٩٥ في مختلف القطاعات الاقتصادية الخدمية في مصر عموما، وفي قطاع التأمين على وجه الخصوص، ببدء تطبيق الإتفاقية العامة لتحرير التجارة في الخدمات General Agreement for Trade of Services والتي تعرف اختصارا بإتفاقية الجاتس GATS. وكان من أهم التغيرات التشريعية التي حدثت في قطاع التأمين في ذلك الوقت هو الغاء النسبة الإلزامية لإعادة التأمين التي كان يُنص عليها في المادة (٣٤) من القانون ١٠ لسنة ١٩٨١ بأن تلتزم جميع شركات التأمين المصرح لها بمزاولة نشاطها فى مصر بإعادة نسبة ٣٠% من جميع عملياتها المقبولة من سوق التأمين المصرى لدى الشركة المصرية لإعادة التأمين، وبالتالي كان الحد الأقصى لاحتفاظ شركة التأمين ٧٠% من جميع عملياتها، ولكن بعد صدور القانون رقم ٩١ لسنة ١٩٩٥ تم الغاء هذه الإلزام، وأصبح لشركات التأمين الحرية في تخطيط سياسة إعادة التأمين وتحديد حجم الاحتفاظ المناسب في ضوء الظروف الخاصة بها، وفي إطار من الضوابط والقواعد التي تضعها هيئة الرقابة على التأمين الخاصة بالسلامة والملاءة المالية للشركة.

وبصفة عامة، يعتبر القرار الخاص بتحديد مستوى الاحتفاظ المناسب من القرارات الهامة والمعقدة في نفس الوقت التي تتخذها شركة التأمين. يرجع تعقد قرار الاحتفاظ إلى تعدد وتداخل المتغيرات المؤثرة على القرار، بالاضافة إلى وجود عدد من المتغيرات التي يصعب قياسها كميا، مما يضطر معها متخذ القرار الاعتماد على الحكم الشخصي بناء على الخبرة وليس بناء على أسس موضوعية. أما بالنسبة أهمية هذا القرار فترجع إلى نواحي عدة منها التأثير المباشر على عدد من القرارت الهامة في شركة التأمين مثل إعادة التأمين، وتخطيط سياسة الاكتتاب، وسياسة الاستثمار، هذا بالإضافة إلى تأثيره على كل من حجم المخاطر التي تتعرض لها شركة التأمين، وعلى حجم الأرباح التي تحققها. بالنسبة للربحية، يرى كل من (Lee K. and *et al*,1992)[١] أنه عندما يكون مستوى الاحتفاظ أقل مما يجب من شأنه أن يحد من قدرة شركة التأمين على الاستخدام والتوظيف الفعال لرأس المال متمثلا في فقد دخل من الاستثمار كان يمكن لشركة التأمين تحقيقه لو كان حجم الاحتفاظ مناسبا، وبالتالي ضياع فرصة لتحقيق عائد

[١] Lee K, Palmer BA, Skipper HD (1992), An Analysis of Life Insurer Retention imits. Journal of Risk and Insurance, 59(1):57-71.

مناسب على رأس المال وزيادة ثقة الممولين في قطاع التأمين أو أو زيادة الأرباح التي يمكن أن تسهم في زيادة نمو شركة التأمين ودعم قيمتها السوقية. أما بالنسبة لدرجة المخاطر، يلاحظ أن الاحتفاظ المبالغ فيه، يمكن أن يترتب عليه تراكم للأخطار داخل محفظة الاكتتابات، وبالتالي زيادة درجة المخاطر مما قد يُعرض شركة التأمين للتعثر وربما للإفلاس خاصة في حالة الحوادث الضخمة. وبالتالي يتعين على شركة التأمين أن تختار المستوى المناسب من الاحتفاظ بالأقساط، بحيث لا يكون أقل من اللازم أو أكثر مما يجب لتفادي الآثار السلبية للمستوى غير المناسب للاحتفاظ.

وبناء على ما سبق، تكون هناك حاجة لوجود أسلوب علمي يمكن لمتخذ القرار في شركة التأمين أن يعتمد عليه في تحديد أهم العوامل والمتغيرات التي يجب أن يأخذها في الاعتبار عند التنبؤ أو تحديد المستوى المتوقع للاحتفاظ في شركة التأمين. خاصة وأن تحديد حد الاحتفاظ في الواقع العملي في الغالب يعتمد إلى حد كبير على قواعد تحكمية[٢].

٢. الدراسات السابقة:

بصفة عامة، يعتبر موضوع الاحتفاظ بالأقساط في شركات التأمين من الموضوعات التي نالت إهتمام كبير من جانب الباحثين سواء داخل مصر أو خارجها. وقد تنوعت هذه الدراسات، فهناك دراسات إهتمت بالتنبؤ بحجم الاحتفاظ (سواء على مستوى الفرع الواحد أو الشركة أو على مستوى سوق التأمين ككل) باستخدام السلاسل الزمنية والتوزيعات الاحتمالية، ودراسات أخرى تناولت العوامل أو المحددات المؤثرة في القرار الخاص بتحديد حجم الاحتفاظ (سواء كانت محددات عامة أو محددات خاصة بفرع أو وثيقة معينة) بالاضافة إلى الدراسات الخاصة بتخطيط الطاقة الاحتفاظية والتوزيع الأمثل لها بين الفروع المختلفة داخل الشركة الواحدة. وهنا سوف يتم التركيز على الدراسات المرتبطة بموضوع البحث الحالي وهي العوامل أو المحددات المؤثرة على قرار الاحتفاظ في شركات التأمين، وهو ومن أمثلة هذه الدراسات:

(أ) دراسة (Gottheimer, 1983) [٣]:

---

[٢] راجع كل من:

- Carter RL (1979), Reinsurance. London: Kluwer Publishing.
- النادي، هويدا جمال الدين، دراسة تحليلية للعوامل المؤثرة في تحديد حدود الاحتفاظ في التأمينات العامة (بالتطبيق على فرع النقل البحري)، رسالة ماجستير غير منشورة، كلية التجارة، جامعة القاهرة، ٢٠٠٤.

[3] Gottheimer, George M ,(1983), A Strategy for Reinsurance Management, CPCU J., 36: 206-218.

قامت هذه الدراسة باختبار مدى وجود علاقة خطية بين حدود الاحتفاظ المحددة من قبل شركات تأمينات الممتلكات والمسؤوليات وكل من: حجم الأصول، الفائض أو الأرباح التي يحصل عليها حملة الأسهم، حجم الأقساط، معدل الخسارة، وذلك بالاعتماد على أسلوب الانحدار الخطي التدريجي stepwise regression. وتوصلت الدراسة إلى أن هناك علاقة خطية موجبة (طردية) بين حدود الاحتفاظ وكل من الأصول وحجم الأقساط.

(ب) دراسة (Jang, 2001) [٤]:

قامت هذه الدراسة بفحص محددات الاحتفاظ لإتفاقيات إعادة تأمين زيادة الخسائر Excess–Of–Loss Reinsurance Retentions في صناعة التأمين في الولايات المتحدة. وتوصلت الدراسة إلى وجود علاقة معنوية موجبة بين مستوى الاحتفاظ بالأقساط وكل من الفائض، ونسبة السيولة، وسعر إعادة التأمين كانت لهم ، في حين أن العلاقة كانت عكسية مع نسبة التشغيل واحتمال التعرض للكوارث Catastrophe Exposure.

(ج) دراسة (Lee-Lee, 2011) [٥]:

كانت هذه الدراسة بالتطبيق على شركات تأمينات الممتلكات في تايوان، وشملت (١٢) متغير، منهم ١٠ متغيرات خاصة بشركة التأمين، واثنان من المتغيرات الاقتصادية العامة. وقد توصلت الدراسة إلى علاقة معنوية موجبة بين حجم الاحتفاظ بالأقساط وكل من: العائد على الأصول، السيولة، العائد على الاستثمار، تركز الأعمال، بينما كانت العلاقة معنوية عكسية مع كل من: مخاطر الاكتتاب، أسعار إعادة التأمين وذلك وفقا لأسلوب الانحدار المبني على طريقة المربعات الصغرى العادية OLS.

ومما هو جدير بالذكر أنه تم الاعتماد على هذه الدراسة في التأصيل النظري لبعض المتغيرات الخاصة بالدراسة الحالية، والفرق بين هذه الدراسة والدراسة الحالية يتمثل في أن الأولى اعتمدت على الطرق المعلمية من خلال أسلوب الانحدار المتعدد باستخدام ٣ نماذج كمية: نموذج انحدار طريقة المربعات الصغرى Ordinary Least Squares Regression، ونموذج

[4] Jang, Jonghag (2001), Reinsurance retentions and limits for property- Liability Insurers: Theory and Empirical Tests. Temple University doctoral Dissertation.

[٥] Lee, Hsu-Hua and Lee, Chen-Ying, Determinants of property-liability insurer retention: Evidence from Taiwan insurance industry, African Journal of Business Management Vol. 5(32), pp. 12543-12550,14 December, 2011

التأثير الثابت Fixed Effects Model ، نموذج التأثير العشوائي Random Effects Model، في حين أن الدراسة الحالية سوف تعتمد على أسلوب لامعلمي وهو أسلوب شجرة القرارات التصنيفية بخوارزمية C4.5 كأحد أساليب تعلم الآلة Machine Learning عند التنقيب عن البيانات Data Mining.

(د) دراسة (Kwangbong and et al, 1992) [٦]

قامت هذه الدراسة ببحث العلاقة بين حدود الاحتفاظ وعدد من المتغيرات المستقلة بالاعتماد على أسلوب الانحدار المبني على المكونات الأساسية The Principal Components Regression Model وقد توصلت الدراسة إلى وجود علاقة معنوية عند مستوى معنوية ١% بين حدود الاحتفاظ في الشركات محل الدراسة وكل من: حجم الشركة، الشكل القانوني للشركة (مساهمة – تبادلي)، متوسط مبلغ التأمين للوثيقة، درجة التركيز على المنتجات الجديدة.

أما بالنسبة للدراسات التي أجريت على مستوى سوق التأمين المصري، نجد أن هناك عدد من الدراسات تناولت تحديد العوامل المؤثرة في تحديد حد الاحتفاظ، بعضها اعتمد على التحليل النظري دون استخدام أسلوب رياضي أو إحصائي معين، مثل دراسة كل من (النادي، هويدا جمال الدين،٢٠٠٤) [٧] ، ودراسة (أحمد، مصطفى عبد الغني، ٢٠٠٠) [٨]. وهناك دراسات أخرى اعتمدت بصفة أساسية على نماذج الانحدار مثل: دراسة (عفيفي، زينب حسن، ١٩٨٨)[٩] التي اعتمدت على نموذج انحدار المكونات الأساسية باستخدام التحليل العاملي، ودراسة (عبد الجواد، عمر أحمد، ١٩٩٩)[١٠] التي استخدمت أسلوب الانحدار المتعدد

---

[6] Lee, Kwangbong and et al, (1992), An analysis of life insurer retention limits, Journal of Risk and Insurance, March, Volume: 59, Issue : 1.

[٧] النادي، هويدا جمال الدين، دراسة تحليلية للعوامل المؤثرة في تحديد حدود الاحتفاظ في التأمينات العامة (بالتطبيق على فرع النقل البحري)، رسالة ماجستير غير منشورة، كلية التجارة، جامعة القاهرة، ٢٠٠٤.

[٨] أحمد، مصطفى عبد الغني، الاتجاهات الحديثة لقياس الطاقة الاستيعابية ومعدل الاحتفاظ، المجلة المصرية للدراسات التجارية، مج ٢٤، ع١، ٢٠٠٠.

[٩] عفيفي، زينب حسن، العوامل المؤثرة في الطاقة الاحتفاظية الصافية بالتأمينات العامة، دراسة عاملية تطبيقية على شركات التأمين المباشر، مجلة الدراسات المالية والتجارية، كلية التجارة – جامعة أسيوط، ١٩٨٨.

[١٠] عبد الجواد، عمر أحمد، إعداد نماذج حدود الاحتفاظ الأمثل باستخدام الأساليب الكمية – دراسة تحليلية بالتطبيق على تأمينات الحياة في السوق المصري، رسالة دكتوراه غير منشورة، كلية التجارة – جامعة المنصورة، ١٩٩٩.

المرحلي، ودراسة (شهاب الدين، محمد مصطفى – مشعال، محمود عبد العال، ٢٠١٣)[١١] التي استخدمت أسلوب الانحدار البوتستراربي، ودراسة (عبد المولى، محمد – المهدي، محمد، ١٩٩٥)[١٢] وقد استخدمت هذه الدراسة أسلوب الانحدار المبني على المكونات الأساسية، ودراسة (إبراهيم، أحمد عبد الرحمن أحمد، ٢٠١٢)[١٣] وقد اعتمدت هذه الدراسة تحليل المسار Path Analysis.

٣. مشكلة البحث:

بناء على ما سبق، وفي ضوء ما تم عرضه من دراسات سابقة، نلاحظ أن معظم الدراسات التي تناولت محددات قرار الاحتفاظ اعتمدت بصفة أساسية على الأساليب المعلمية Parametric Techniques عموما وعلى نماذج الانحدار المتعدد على وجه الخصوص، وهي أساليب يرتبط تطبيقها بتوافر عدد كبير من الشروط النظرية، بعضها قد يصعب توافرها في الواقع العملي خاصة تلك المتعلقة بالتكامل المشترك لتفادي الوقوع في مشكلة الانحدار الزائف Spurious Regression. كما أن جميع الأبحاث التي شملتها الدراسة لم تقم بالتأكد من توافر الشروط التي تتطلبها هذه النماذج مما يقلل من درجة الاعتمادية والثقة في النتائج التي تم التوصل اليها، الأمر الذي يجعلنا في حاجة إلى البحث عن أساليب رياضية أخرى لا تتطلب نفس القدر من الشروط وتكون سهلة الفهم والتطبيق في الحياة العملية. وبالتالي تكون مشكلة البحث تتمثل في الحاجة الى نموذج رياضي مبسط يسهل فهمه وتطبيقه من جانب متخذ القرار في شركات التأمين وفي هيئات الرقابة على التأمين من أجل التنبؤ بمستوى الاحتفاظ بالأقساط.

---

[١١] شهاب الدين، محمد مصطفى – مشعال، محمود عبد العال، استخدام أسلوب البوتستراربي في تقدير النموذج الأمثل لدالة الاحتفاظ بالأقساط في سوق التأمين السعودي، مجلة البحوث المعاصرة، كلية التجارة – جامعة سوهاج، ٢٠١٣

[١٢] عبد المولى، محمد – المهدي، محمد، نحو إطار متكامل لتحديد العوامل المؤثرة على حدود الاحتفاظ في التأمين على الحياة، دراسة تطبيقية على شركات التأمين المباشر العربية، مجلة آفاق جديدة، كلية التجارة – جامعة المنوفية،العدد الثاني، ١٩٩٥.

[١٣] إبراهيم، أحمد عبد الرحمن أحمد، استخدام نموذج تحليل المسار في تحديد أهم العوامل المؤثرة على معدل الاحتفاظ لشركات التأمين المصرية –دراسة تطبيقية، مجلة البحوث المالية والتجارية، كلية التجارة – جامعة بورسعيد، العدد الثاني، ٢٠١٢.

٤. أهداف البحث: تتمثل أهم اهداف البحث في :

(أ) دراسة وتحليل أهم محددات الاحتفاظ بالأقساط في تأمينات الممتلكات والمسؤولية في سوق التأمين المصري.

(ب) بناء نموذج رياضي بالاعتماد على شجرة القرارات التصنيفية المبنية بخوارزمية C4.5 في تعلم الألة Machine Learning للتنبؤ بمستوى الاحتفاظ بالأقساط في تأمينات الممتلكات والمسؤولية في سوق التأمين المصري، في ضوء المحددات التأمينية والمحددات الاقتصادية العامة في السوق.

٥. أهمية البحث:

٥.١ من الناحية الأكاديمية:

تتمثل أهم المساهمات الأكاديمية في أن هذا البحث ربما يكون هو الأول سواء المستوى المحلي أو الدولي – على حد علم الباحث– الذي يتم الاستعانة بأساليب التنقيب عن البيانات Data Mining عموما وأسلوب شجرة القرارات التصنيفية على وجه الخصوص من أجل التنبؤ بمستوى الاحتفاظ بالأقساط، وبالتالي فهو يساعد على سد الفجوة البحثية في هذه الجزئية، وهو يعد إثراء لأدبيات التأمين في مجال الاحتفاظ بالأقساط.

٥.٢ من الناحية التطبيقية:

تتمثل الأهمية العملية أو التطبيقية لهذا البحث في توفير أداة تعتمد على أساس علمي يمكن أن يعتمد عليها متخذ القرار سواء في شركات التأمين أو هيئات الرقابة على التأمين وهو بصدد التنبؤ بمستوى الاحتفاظ بالأقساط في شركة التأمين أو في سوق التأمين في ضوء المحددات التي سوف تتناولها الدراسة.

٦. متغيرات الدراسة:

عند بناء شجرة القرارات التصنيفية Classification Decision Tree يوجد نوعين من المكونات، المكون الأول ويعرف بالخصائص Attributes وهي تمثل المتغيرات المستقلة، أما المكون الثاني فيعرف بالمخرجات Output وهو يمثل المتغير التابع.

أولا: المخرجات (المتغير التابع):

تتمثل المخرجات (المتغير التابع) في نموذج شجرة القرارات التصنيفية في مستوى الاحتفاظ بالأقساط في سوق تأمينات الممتلكات والمسؤولية في مصر. ونظرا لأن الشرط الأساسي لشجرة القرارات التصنيفية أن تكون المخرجات متغير وصفي فئوي Categorical Variable، سيتم وضع حدود

تحكمية Rule of Thumb لمستويات الاحتفاظ، تعتمد على الحد الأدنى والحد الأعلى لنسب الاحتفاظ بالأقساط التي يتم حسابها خلال فترة الدراسة كما يلي:

نسب الاحتفاظ بالأقساط = صافي الأقساط المكتتبة ÷ إجمالي الأقساط المكتتبة

حيث أن:

صافي الأقساط المكتتبة = إجمالي الأقساط المكتتبة – أقساط إعادة التأمين الصادر (المحلي والخارجي).

وإجمالي الأقساط المكتتبة = أقساط الاكتتاب المباشر + أقساط إعادة التأمين الوارد (المحلي والخارجي).

٦.٢ الخصائص (المتغيرات المستقلة):

تتمثل الخصائص في المتغيرات المستقلة التي من المتوقع أن يكون لها تأثير على التنبؤ بمستوى الاحتفاظ في سوق التأمين من وجهة نظر الباحث. وقد تم هذه الخصائص أو المتغيرات الى نوعين أساسيين هما:

أولا: المتغيرات التأمينية:

وتضم هذه المجموعة بصفة أساسية المتغيرات المتعلقة بالنواحي الفنية الخاصة بشركات التأمين وسوق التأمين عموما. وتتمثل أهم هذه المتغيرات من وجهة نظر الباحث وفي ضوء ما توصلت اليه الدراسات السابقة في:

(١) حجم شركات التأمين:

في دراسة قام كل من Kwangbong وآخرون[١٤] شملت ٩٧ شركة تأمين حياة في الولايات المتحدة الأمريكية، لمعرفة مدى وجود علاقة بين مستوى الاحتفاظ بالأقساط وعدد من المتغيرات من بينها حجم شركة التأمين، تبين أن حجم شركة التأمين هو أكثر المتغيرات أهمية من حيث التأثير على حدود الاحتفاظ في شركات التأمين محل الدراسة.

ويرجع السبب وراء قدرة شركات التأمين الكبيرة على الاحتفاظ بالأقساط بدرجة أكبر من الشركات الصغيرة من وجهة نظر كل من Cummins and Sommer[١٥] إلى أن الشركات الكبيرة تمتلك محفظة اكتتابات كبيرة ومنوعة تنوعيا جيدا يساعدها على تحقيق قانون الأعداد

[١٤] Lee, Kwangbong and et al, (1992), Op. cit.

[١٥] Cummins, David, Sommer, David, (1996), Capital and risk in property-liability insurance markets. Journal of Banking & Finance, vol. 20, issue 6, 1069-1092.

الكبيرة مما يقلل من خطر الانحرافات بين النتائج الفعلية والنتائج المتوقعة في المحفظة عند أدنى حد ممكن بالمقارنة بالشركات صغيرة الحجم ، ومن ثم تستطيع الاحتفاظ بالمزيد من الأقساط.

وفي تفسير آخر لنفس العلاقة، يرى كل من Mayers and Smith [١٦] أن قدرة شركة التأمين على تحقيق زيادة في التدفقات النقدية الداخلة سواء من الأقساط أو عائد استثمارها، يرتبط بشكل مباشر بحجم الشركة وما تمتلكه من أصول، بافتراض الاستخدام الفعال للموارد المادية والبشرية التي لديها.

وأيضا فى دراسة لكل من Hammond and Shapiro [١٧] حول العلاقة بين الخطر في محفظة الاكتتاب (معبرا عنه بالانحراف المعيارى لنتائج الاكتتاب) وحجم رأس المال، توصلت الدراسة إلى أن هناك علاقة عكسية بين المتغيرين، فكان الانحراف المعياري لنتائج الإكتتاب في شركات التأمين الصغيرة والشركات الجديدة أكبر من الانحراف المعياري في الشركات كبيرة الحجم. وقد أوضحت الدراسة أن السبب في ذلك يرجع إلى أن الشركات الكبيرة تستطيع تحقيق معدل مصروفات مستقر بسبب حجم الوفورات التي تحققها، بالاضافة إلى أن معدل الخسارة في الشركات الكبيرة يتسم بإستقرار نسبي وذلك بالمقارنة بالشركات الصغيرة.

وبناء عليه، من المحتمل أن يكون هناك علاقة طردية بين حجم شركات التأمين وحجم الاحتفاظ بالأقساط في سوق التأمين المصري.

وفيما يتعلق بالقياس الكمي لمتغير حجم سوق التأمين، نجد أن أكثر ٣ مقاييس تم الاعتماد عليهم في قياس حجم الشركات في آدبيات التمويل عموما هي: إجمالي الأصول، حجم المبيعات، القيمة السوقية للشركة. وفي دراسة مسحية Survey Study قام بها كل من Dang and Li [١٨] شملت ١٠٠ بحث تجريبي (دراسة تطبيقية) Empirical Study في كبرى المجلات العلمية في مجال التمويل عن أهم المقاييس المستخدمة في قياس حجم الشركة، وجدت

---

[١٦] Mayers, David, Smith, Clifford., (1990), On the Corporate Demand for Insurance: Evidence from Reinsurance Market. Journal of Business 63 (1), 19:40.

[١٧] Hammond, J. D., Shapiro, Arnold F.,(1979), Capital Requirements for Entry into Property and Liability Underwriting: An Empirical Examination, , APR 75-16550 A01, P.38.

[١٨] *Dang, Chongyu , and Li, Frank , (2015), Measuring Firm Size in Empirical Corporate Finance, Journal of Banking & Finance Volume 86, January 2018, Pages 159-176.*

أن ٨٧ دراسة اعتمدت على المقاييس الثلاثة، وهذه الأبحاث موزعة بالشكل التالي: ٤٩ دراسة استخدمت إجمالي الأصول، في حين أن ٢٠ دراسة اعتمدت على القيمة الرأسمالية أو القيمة السوقية، بينما ١٦ دراسة استخدمت حجم المبيعات، ودراستان فقط استخدمت عدد العاملين بالشركة كمقياس لحجم الشركة، وفي جميع الأبحاث تم حساب اللوغاريتم الطبيعي لكل مقياس من هذه المقاييس الثلاثة. وبالتالي سوف يعتمد الباحث في الدراسة الحالية على اللوغاريتم الطبيعي لإجمالي أصول شركة التأمين لقياس المتغير الخاص بحجم سوق التأمين.

(٢) الرافعة المالية لسوق التأمين:

بصفة عامة، تقيس الرافعة المالية مخاطر التمويل في شركة التأمين. ويشير هنا كل من Lee and Lee [١٩] أنه العلاقة بين الرافعة المالية لشركة التأمين ومستوى الاحتفاظ علاقة عكسية، فكلما كانت رافعة التأمين Insurance leverage كبيرة، انخفض احتفاظ شركة التأمين، والعكس بالعكس. وفيما يتعلق بالقياس الكمي لهذا المتغير، يلاحظ أنه يوجد في مجال التأمين، يوجد نوعين أساسيين من المقاييس التي تُستخدم في حساب الرافعة المالية لشركة التأمين وهما:

(أ) نسبة صافي الأقساط المكتتبة إلى الفائض (حقوق المساهمين) Net Premiums Written to Policyholder Surplus Ratio.

(ب) نسبة المخصصات الفنية إلى الفائض (حقوق المساهمين).

ويعتبر المقياس الأول هو الأكثر انتشارا واستخداما في أدبيات التأمين، فضلا عن أنه أحد المقاييس التي تعتمد عليها IRIS، كما أنه يتسم بالبساطة وسهولة الاستخدام[٢٠]. وبناء عليه في الدراسة الحالية سوف يتم الاعتماد على المقياس الأول عند حساب الرافعة المالية لسوق التأمين.

(٣) مخاطر محفظة إكتتابات محفظة السوق:

هنا يرى Sommer D [٢١] أن هناك علاقة عكسية بين مخاطر الاكتتاب ومستوى الاحتفاظ بالأقساط، فكلما كان حجم الخسائر المتوقعة في محفظة الإكتتابات كبيرا، انخفض حجم

---

[١٩] Lee, Hsu-Hua and Lee, Chen-Ying, Op. cit.

[٢٠] Szczepanski, Chester J., Cooper, Craig A., An Analysis of Variations in Leverage Ratios Among Insurers, https://www.casact.org/pubs/dpp/dpp92/92dpp269.pdf.

21 Sommer D. (1996), The Impact of Firm Risk on Property-liability Insurance Price. J. Risk Insur., 63(3):501- 514.

الاحتفاظ في محفظة الاكتتاب من خلال نقل الخطر لمعيد التأمين، بعكس الحال لو أن حجم الخسائر المتوقعة في محفظة الإكتتابات كان منخفضا، سيكون احتفاظ شركة التأمين أكبر من أجل تحقيق المزيد من الأرباح.

في الدراسة الحالية سوف يعتمد الباحث على المنهج الذي اتبعه كل من Adams and Buckle لقياس مخاطر الاكتتاب من خلال الاعتماد على معدل الخسارة في سوق التأمين في السنة السابقة أي عند الفترة (t−1) [٢٢].

(٤) تكلفة إتفاقيات إعادة التأمين:

يرى David Cummins وآخرون[٢٣] أن هناك علاقة عكسية بين تكلفة اتفاقيات اعادة التأمين ومستوى الاحتفاظ بالأقساط في سوق التأمين، حيث أن إرتفاع تكلفة إعادة التأمين من شأنه أن يقلل من الطلب على إعادة التأمين، وبالتالي زيادة حجم الأقساط المحتفظ بها في شركات التأمين.

وفيما يتعلق بقياس هذا المتغير، نجد أن كل من Lee and Lee [٢٤] قد اعتمدا على نسبة أقساط إعادة التأمين مطروحا منها عمولة إعادة التأمين المكتسبة إلى المطالبة المستردة من معيدي التأمين عند حساب تكلفة إتفاقيات إعادة التأمين. وفي الدراسة الحالية نظرا لعدم توافر بيانات منشورة في سوق التأمين المصري عن عمولة إعادة التأمين سوف يتم قياس هذا المتغير من خلال قسمة أقساط إعادة التأمين الصادر على المطالبة المستردة من معيدي التأمين (مقلوب معدل الخسارة لعمليات إعادة التأمين الصادر).

(٥) معدل نمو الأقساط وحجم الأعمال:

يرى كل من Black and Skipper [٢٥] أن كلما كانت شركة التأمين تحقق معدلات نمو متزايدة في حجم أعمالها، وبالتالي زيادة مضطردة في حجم الأقساط والتدفقات النقدية الداخلة، أمكنها زيادة حدود احتفاظها من هذه الأقساط. أي أن هناك علاقة طردية بين نمو الأقساط

---

[٢٢] Adams, M., Buckle, M., (2003),The determinants of corporate financial performance in the Bermuda insurance market. Appl. Finance. Econ., 13: 133-143.

[٢٣] Cummins JD, Dionne G Gagné R, Nouira A, Song QF (2008), The Costs and Benefits of Reinsurance. Working Paper, SCOR/JRI conference.

[٢٤] Lee, Hsu-Hua and Lee, Chen-Ying, Determinants of property-liability insurer retention: Evidence from Taiwan insurance industry, African Journal of Business Management Vol. 5(32), pp. 12543-12550,14 December, 2011

[٢٥] Black K, Skipper HD (1994), Life Insurance. 12th ed., Prentice Hall, New York.

وحجم الاحتفاظ في سوق التأمين. ويرى الباحث هنا أن شكل العلاقة بين حجم الأعمال ممثلة في حجم الأقساط ومستوى الاحتفاظ يتوقف على عوامل أخرى أهمها السيولة ومدى حاجة شركات التأمين إلى التدفقات النقدية الداخلة لسداد لتزاماتها، فإذا كانت شركات التأمين تحقق معدلات نمو متزايدة من الأقساط وفي ظل توافر السيولة الكافية في محفظة استثماراتها يمكنها التقليل من مستوى الاحتفاظ، بعكس الحال لو أن هذه الشركات لا يتوافر لديها السيولة الكافية فمن المؤكد في هذه الحالة سيكون القرار هو الاحتفاظ بالمزيد من الأقساط.

وفيما يتعلق بقياس هذا المتغير، سوف تعتمد الدراسة الحالية على الأسلوب الذي استخدمه كل من Cummins and Nini[٢٦] من خلال قسمة إجمالي الأقساط المكتتبة للسنة الحالية (t) على إجمالي الأقساط المكتتبة السنة السابقة (t−1). وغني عن البيان، كما سبق الاشارة اليه، أن إجمالي الأقساط المكتتبة تشمل أقساط الاكتتاب المباشر مضافا اليها أقساط إعادة التأمين الوارد المحلي والخارجي.

(٦) نسبة السيولة:

تقيس نسبة السيولة قدرة شركة التأمين على الوفاء بالتزاماتها تجاه حملة الوثائق والدائنين. وبصفة عامة، تشير الدراسات المالية وإدارة المخاطر المالية إلى أنه كلما كانت نسبة السيولة كبيرة، قلت احتمالية حدوث أزمة مالية. ويرى كل من Lee and Lee [٢٧] أنه عندما تمتلك شركات التأمين سيولة كافية يكون لديها القدرة على الاحتفاظ بالمزيد بالأقساط مقارنة بالشركات التي لا يتوافر لديها نفس القدر من السيولة. وهنا يرى الباحث أن العكس هو الصحيح، وأن العلاقة بين مستوى الاحتفاظ بالأقساط وحجم السيولة المتوافرة لدى شركات التأمين هي علاقة عكسية وليست طردية، حيث أنه يكون من المنطقي كلما واجهت شركة التأمين مشكلة في السيولة زاد حجم احتفاظها بالأقساط والاستفادة من عائد استثمارها في توفير الأموال اللازمة لسداد التزاماتها. وفي الدراسة الحالية سيتم قياس السيولة بالاعتماد على الأسلوب الذي

[٢٦] Cummins JD, Nini G P (2002), Optimal Capital Utilization by Financial Firms: Evidence from the Property-Liability Insurance Industry. J. Financ. Serv. Res., 21(1): 15-53.

[٢٧] Lee, Hsu-Hua and Lee, Chen-Ying Op. cit.

استخدمه كل من Chen and Wong من خلال قسمة الأصول السائلة على إجمالي الالتزامات [٢٨].

(٧) معدل العائد على الاستثمار:

مما لا شك فيه أن كلما كانت شركات التأمين قادرة على تحقيق معدلات مرتفعة من العائد على الاستثمار، تستطيع أن تحتفظ بقدر أكبر من الأقساط مقارنة بالحالة التي يكون فيها معدلات العائد على الاستثمار التي تحققها الشركات منخفضة. وبالتالي فإن العلاقة المتوقعة بين مستوى الاحتفاظ بالأقساط في سوق التأمين ومعدل العائد على الاستثمار علاقة طردية[٢٩]. وغني عن البيان أن معدل العائد على الاستثمار يتم حسابه من قسمة صافي الدخل من الاستثمار على إجمالي الأصول المستثمرة.

(٨) خبرة سوق التأمين:

كلما كانت خبرة سوق التأمين بطبيعة وسلوك الأخطار التي تقوم بتغطيتها طويلة، أمكن التنبؤ بدقة بإحتمال وحجم الخسارة المتوقعة، وبالتالي التقدير الدقيق لدرجة الخطورة في محفظة الاكتتابات، وهذا من شأنه أن يساعد على زيادة الاحتفاظ بالأقساط مقارنة بالأسواق حديثة العهد بالتأمين التي لا يتوافر لديها الخبرة الكافية أو القدرة على التقدير الدقيق لدرجة الخطورة. وبالتالي تكون هناك علاقة طردية بين خبرة سوق التأمين وحجم الاحتفاظ بالأقساط. وفي الدراسة الحالية سوف يتم الاعتماد على متغير الزمن لقياس خبرة سوق التأمين.

(٩) عدد شركات التأمين العاملة في سوق التأمين:

تزيد قدرة أسواق التأمين على الاحتفاظ بالمزيد من الأقساط في حالتين: حالة الأسواق الاحتكارية أو شبه الاحتكارية، وحالة الأسواق غير الاحتكارية والمنافسة غير سعرية. وبالتالي كلما زاد عدد شركات التأمين العاملة في سوق التأمين، تزيد قدرة سوق التأمين على الاحتفاظ بالأقساط بسبب زيادة الخبرة والمنافسة ورغبة الشركات في تحقيق النمو. أما في حالة المنافسة السعرية فإن زيادة عدد شركات التأمين العاملة في سوق التأمين وتحت ضغط المنافسة (خاصة المنافسة السعرية) قد تضطر معه شركات التأمين الى اللجؤ لسياسة اكتتاب متساهلة عند قبول

[٢٨] Chen R, Wong KA (2004),The Determinants of Financial health of Asian Insurance Companies. J. Risk Insur., 71(3): 469-499.

[٢٩] Gatzlaff K (2009),Dimensions of Property- Liability insurer Performance. The Florida State University doctoral Dissertation.

الأخطار ومن بينها تقليل سعر التغطيات التأمينية، الأمر الذي قد يترتب عليه قبول أخطار رديئة أو مرتفعة الخطورة نوعا ما، مما يزيد من درجة الخطورة في محفظة الاكتتابات لديها، وهذا من شأنه أن يقلل من قدرة الشركات على الاحتفاظ بالأقساط. وبالتالي يمكن القول بأن هناك علاقة عكسية بين مستوى الاحتفاظ بالأقساط وحدة المنافسة في سوق التأمين. في الدراسة الحالية سوف يتم قياس هذا المتغير من خلال عدد شركات التأمين العاملة في سوق التأمين في السنة السابقة (t−1) خلال فترة الدراسة.

ثانيا: المتغيرات الاقتصادية:

تضم هذه المجموعة المتغيرات الاقتصادية العامة، وتتمثل أهم هذه المتغيرات من وجهة نظر الباحث وفي ضوء ما توصلت اليه الدراسات السابقة في:

(١٠) التضخم:

بصفة عامة، يتم احتساب الأقساط الخاصة بمحفظة التأمين بناءً على تقدير معين للخسائر والتكاليف المستقبلية في ظل توقع لمعدل تضخم معين. وبالتالي لو أن معدل التضخم الفعلي كان أكبر من المتوقع، فمن المحتمل أن يحد أو يقلل من قدرة شركات التأمين على الوفاء بالتزاماتها وقت الاستحقاق، وهذا من شأنه أن يزيد من إحتمال التعثر المالي أو على الأقل التأثير على هامش الربح في منشأة التأمين[٣٠]. وبالتالي من المفترض أن تكون العلاقة بين معدل التضخم المتوقع ومستوى الاحتفاظ بالأقساط علاقة عكسية، فكلما كان معدل التضخم المتوقع كبيرا، قل مستوى الاحتفاظ بالأقساط في شركات التأمين، والعكس بالعكس.

(١١) معدل النمو الاقتصادي:

في دراسة لكل من Grace and Hotchkiss [٣١] أوضحت أن هناك علاقة طويلة الأجل بين الناتج المحلي الإجمالي ونتائج محفظة الإكتتاب، وأن هناك علاقة طردية بين معدل النمو الاقتصادي وحجم الاحتفاظ بالأقساط في سوق التأمين، حيث أن زيادة الناتج القومي الإجمالي يترتب عليها زيادة في الدخل القومي الفعلي، في الغالب يتبعها زيادة في طلب التأمين وبالتالي زيادة في حجم الأقساط، الأمر الذي ومن المحتمل زيادة في حجم الأقساط المحتفظ بها. ويرى

[30] Browne MJ, Hoyt RE (1995),Economics and Market Predictors and Insolvencies in the Property-Liability insurance Industry. J. Risk Insur., 62(2): 309-327.

[٣١] Grace, Martin F. and Hotchkiss, Julie L., External Impacts on the Property-Liability Insurance Cycle, Journal of Risk and Insurance 62 (December 1995): 738-54., 2011

الباحث أن هناك سبب آخر وراء زيادة الاحتفاظ بالأقساط في أسواق التأمين المصاحبة للنمو الاقتصادي المرتفع وهو رغبة شركة التأمين في الاستفادة من معدلات الاستثمار المرتفعة التي تصاحب معدلات النمو الاقتصادي المرتفع.

(١٢) أسعار سعر الصرف:

يرى كل من القاضي وآخرون [٣٢] أنه غالبا ما يعزف معيد التأمين عن العمل في أسواق التأمين التى يتسم فيها سعر الصرف بعدم الإستقرار، وبالتإلى نجد أن حجم الاحتفاظ بالأقساط من جانب شركات التأمين يكون أكبر فى حالة أسواق التأمين التى يتسم فيها سعر الصرف بالتذبذب، وذلك بالمقارنه بالاسواق التأمينيه التى يتسم فيها بالثبات النسبي. كما يرى أيضا كل من عبد الباري وحسن [٣٣] أن تاثير التغير في سعر الصرف علي الطاقة الاحتفاظية لأسواق التأمين يكون منخفضا في الحاله التى يكون فيها التغير في سعر الصرف ظاهره عامه في أسواق اعادة التأمين العالمية.

٦. شجرة القرارات التصنيفية وفلسفة عمل خوارزمية C4.5 :

بصفة عامة يعتبر التنقيب عن البيانات Data Mining (DM) أداة مفيدة لاكتشاف المعلومات والمعرفة من البيانات الضخمة Big Data، ويوجد العديد من الأساليب التي يتم الاعتماد عليها في عملية التنقيب. وبصفة عامة يوجد خمسة أنواع أساسية من عمليات DM: التصنيف Classification، الانحدار Regression، تحليل الرابط Link Analysis، ، البحث عن الانحراف Deviation Detection. بحيث يتم الاعتماد على كل من التصنيف Classification، الانحدار Regression في عملية التنبؤ، بينما يستخدم باقي الأساليب في عملية التوصيف لأنماط البيانات Description of Patterns in the Data [٣٤].

من أشهر الأدوات المستخدمة في عملية التصنيف أداة شجرة القرار Decision Tree[٣٥]، ويوجد العديد من الخوارزميات المستخدمة في بناء شجرة القرارات، من أهم هذه الخوارزميات:

---

[٣٢] القاضي، عبد الحليم عبد الله، وآخرون، دراسة الطاقة الاستيعابية لسوق التأمين، أكاديمية البحث العلمي، ٢٠٠٤.

[٣٣] عبد البارئ، محمد وحيد، حسن، نادية احمد، استخدام الاساليب الكمية لتحديد جد الاحتفاظ فى تأمينات الممتلكات والمسئولية المدنية تجاه الغير بالتطبيق على فرع الحريق المجلة المصرية للدراسات التجارية كلية التجارة- جامعة المنصورة – العدد الثانى – الجزء الثانى، ١٩٩٩.

[٣٤] Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.

[35] Badr Hssina and *et al*, A comparative study of decision tree ID3 and C4.5, International Journal of Advanced Computer Science and Applications, Special Issue

ID3, C4.5, CART، وتعتبر خوارزمية C4.5 التي قدمها Ross Quinlan عام ١٩٩٣[٣٦] أكثرهم انتشارا واستخداما وفاعلية وأفضلهم نظرا لما تتمتع به من دقة وسرعة في الوصول للحل [٣٧] ، ومما هو جدير بالذكر أن خوارزمية C4.5 قد تضمنت تعديلات وتحسينات كبيرة في خوارزمية ID3 التي قدمها أيضا نفس العالم سنة ١٩٨٦[٣٨].

تعتمد خوارزمية C4.5 كأسلوب من أساليب تعلم الألة على عملية التدريب، من خلال تقسيم البيانات إلى مجموعتين؛ المجموعة الأولى تسمى بمجموعة التدريب Training set يتم الاعتماد عليها في بناء النموذج من خلال تدريب الألة على عملية التصنيف والمجموعة الثانية تسمى مجموعة الاختبار Test Set والهدف منها معرفة مدى دقة النموذج وقدرته على التفريق بين الحالات المختلفة للمخرجات وبالتالي مدى إمكانية الاعتماد عليه في التنبؤ بالظاهرة محل الدراسة.

٦.١ مفاهيم أساسية:

يعتمد بناء شجرة القرارات التصنيفية على مقياسين أساسيين في نظرية المعلومات:

١ – **مقياس Entropy** : يعد هذا المقياس أحد أهم المقاييس المستخدمة في نظرية المعلومات Information theory التي يتم الاعتماد عليها في قياس عدم التأكد أو حجم المعلومات الخاصة بمتغير عشوائي ما[٣٩]، ويعتبر أول من قدم مصطلح Information Entropy هو العالم Claude Shannon[٤٠]. وتتراوح قيمة هذا المقياس بين الصفر والواحد الصحيح، وكما هو موضح بالشكل التالي، ويلاحظ هنا أنه كلما اقترب قيمة الاحتمال من ٠.٥ كلما كان قيمة مقياس Entropy تقترب من الواحد الصحيح[٤١].

---

on Advances in Vehicular Ad Hoc Networking and Applications. www.ijacsa.thesai.org

[٣٦] Guo, Ligia, Applying Data Mining Techniques in Property and Causality Insurance, https://www.casact.org/pubs/forum/03wforum/03wf001.pdf.

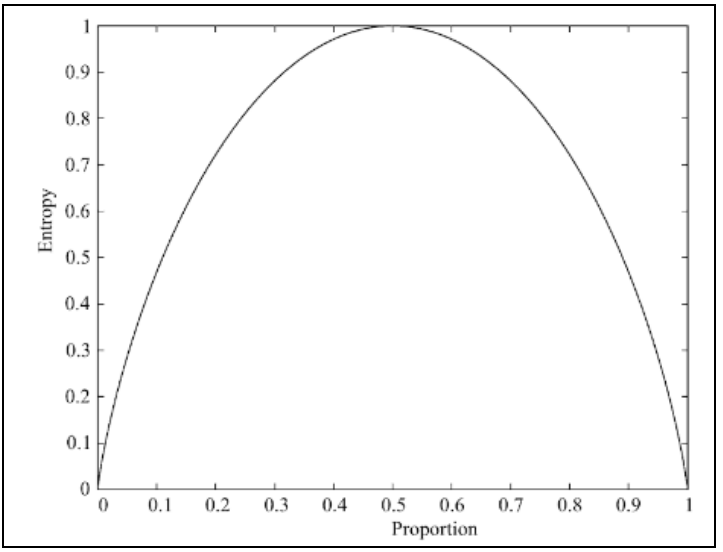
[٣٧] Sharma, Seema and et al, Classification Through Machine Learning Technique: C4.5 C4.5 Algorithm based on Various Entropies, International Journal of Computer Applications, Volume 82 – No 16, November 2013.

[٣٨] Quinlan, J. R. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (Mar. 1986), 81–106.

[٣٩] Max Bramer, Principles of Data Mining, 3rd edition, Springer, Nov 9, 2016.

[٤٠] Shannon, Claude E. (July–October 1948), "A Mathematical Theory of Communication", Bell System Technical Journal, 27 (3): 379–423. http://cm.bell-labs.com/cm/ms/what/shannday/paper.html

[41] Zengchang Qin, Yongchuan Tang, Uncertainty Modeling for Data Mining: A Label Semantics Approach, Springer, 2014.



الشكل (١)

مقياس Entropy

وعند حساب مقياس Entropy يتم التفرقة بين حالتين :

(أ) في حالة التصنيف الثنائي Binary Classification:

$$Entropy\,(S) = -(P^{+}).Log_2(P^{+}) - (P^{-}).Log_2(P^{-})$$

حيث:

P^{+} : تمثل الاحتمال الخاص بالحالة الأولى للمتغير العشوائي.

P^{-}: تمثل الاحتمال الخاص بالحالة الثانية للمتغير العشوائي.

غني عن البيان أن:

$$P^{+} + P^{-} = 1$$

(ب) في حالة التصنيف المتعدد N-ary Classification:

$$Entropy\,(S) = \sum_{i}^{n} - P_i\ Log_2\ P_i$$

ولمزيد من التوضيح لمفهوم مقياس Entropy، سنفرض أننا لدينا متغيرين عشوائيين (X1, X2) كما هو موضح بالجدول التالي، وبحساب مقياس Entropy لكل منهما، نجد أن قيمته في حالة المتغير (X2) أكبر من المتغير (X1)، وأن قيمته في حالة (X2) تقترب من الواحد الصحيح فهو دليل على أن المشاهدات تقترب من التوزيع بالتساوي.

X	A	A	B	B	B	B	B	B	B	B
X	A	A	A	A	B	B	B	B	B	B

$$Entropy\ (X1) = -\frac{2}{10}.Log_2\frac{2}{10} - \frac{8}{10}.Log_2\frac{8}{10} = 0.72$$

$$Entropy\ (X2) = -\frac{4}{10}.Log_2\frac{4}{10} - \frac{6}{10}.Log_2\frac{6}{10} = 0.91$$

وبالتالي يكون عدم التأكد في حالة المتغير (X2) أكبر من (X1). وغني عن البيان، أن عدم التأكد يكون عند القيمة القصوى له (قيمة Entropy تساوي الواحد الصحيح) عندما يكون الاحتمال ٠.٥٠ في حالة التجارب أو المتغيرات العشوائية ذات الحدين Binary Random Variables، أما في حالة N–ary Random Variables وهي الحالة التي يأخذ فيها المتغير(n) من النتائج Outcomes، فإن عدم التأكد عند القيمة القصوى في حالة أن يكون الاحتمال الخاص بكل نتيجة يساوي $(\frac{1}{n})$.

٢ – **مقياس Information Gain** : يعد الهدف من مقياس Information Gain هو حساب حجم التخفيض المتوقع في عدم التاكد (مقدار التخفيض في قيمة Entropy) عند تقسيم أو تصنيف المشاهدات او الحالات وفقا لخاصة ما Attribute (متغير مستقل)، وبالتالي نستطيع قياس مدى قدرة خاصية ما على أفصل تصنيف للحالات الخاصة بالمتغير التابع[٤٢].

وبصفة عامة، عند بناء شجرة القرارات التصنيفية، كلما كان قيمة هذا المقياس كبيرة لأحد المتغيرات، كان هذا المتغير أكثر فاعلية في عملية التصنيف. ويتم حساب قيمة هذا المقياس وفقا للعلاقة التالية :[٤٣]

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|Sv|}{|S|} Entropy(Sv)$$

[42] Lior Rokach and Oded Maimon, Data Mining with Decision Trees: Theory and Applications, World Scientific Publishing, 2008.

[43] Chauhan, Harvinder, and Chauhan, Anu Implementation of decision tree algorithm C4.5, International Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013.

٦.٢ مثال توضيحي:

لتوضيح كيفية بناء شجرة القرارات التصنيفية بالتطبيق على مشكلة مستوى الاحتفاظ بالأقساط في شركات التأمين، سوف يقدم الباحث المثال الافتراضي التالي لعدد ١٤ شركة تأمين تعمل في مجال تأمين الحريق في أحد أسواق التأمين، كما يلي [٤٤] :

جدول رقم (١)
بيانات المثال التوضيحي

الشركة	خبرة شركة التأمين (X1)	معدل العائد على الاستثمار (X2)	تكلفة إتفاقيات إعادة التأمين (X3)	مخاطر الاكتتاب (X4)	مستوى الاحتفاظ (Y)
1	40	معدل عائد مرتفع	1.3	0.95	0.30
2	50	معدل عائد منخفض	1.6	1.05	0.15
3	35	معدل عائد مرتفع	3.4	0.30	0.65
4	8	معدل عائد مرتفع	0.6	1.4	0.70
5	5	معدل عائد منخفض	0.4	0.40	0.80
6	15	معدل عائد مرتفع	2.5	1.1	0.25
7	3	معدل عائد مرتفع	0.9	1	0.55
8	20	معدل عائد مرتفع	0.8	1.9	0.60
9	25	معدل عائد منخفض	0.3	1.08	0.85
10	28	معدل عائد منخفض	2.6	0.60	0.55
11	50	معدل عائد مرتفع	0.5	0.70	0.90
12	30-10	معدل عائد منخفض	1.9	2.1	0.30
13	4	معدل عائد منخفض	0.2	3.5	0.35
14	22	معدل عائد منخفض	1.7	1.7	0.75

هنا يتم تحويل المتغيرات السابقة إلى متغيرات فئوية Categorical Variables كما يلي:

[٤٤] تم توليد بيانات هذا المثال من خلال محاكاة المثال الخاص بلعب رياضة التنس الموجود في المرجع التالي:
Qin, Zengchang, and Tang, Yongchuan, (2014), Uncertainty Modeling for Data Mining: A Label Semantics Approach, Springer, P.78.

جدول رقم (٢)
المستويات المختلفة للمتغيرات الخاصة بالمثال التوضيحي

المتغير	وصف الحالات
Y	(-Ret): احتفاظ منخفض، نسبة الاحتفاظ أقل من أو يساوي 40%. (+Ret): احتفاظ مرتفع، نسبة الاحتفاظ أكبر من 40%.
X1	Exp1: عدد سنوات الخبرة أقل من 10 سنوات. Exp2: عدد سنوات الخبرة من 10 سنوات الى أقل من 30 سنة Exp3: عدد سنوات الخبرة 30 سنة فأكثر.
X2	RA: معدل عائد مرتفع. RT: معدل عائد منخفض.
X3	(-ReC): تكلفة منخفضة لإتفاقيات إعادة التأمين، أقل من 100%. (+ReC): تكلفة مرتفعة لإتفاقيات إعادة التأكثر 100% فأكثر.
X4	(LR1): مخاطر اكتتاب منخفضة، معدل الخسارة أقل من 0.90 . (LR2): مخاطر اكتتاب متوسطة، معدل الخسارة من 0.90 الى أقل من 1.20 . (LR3): مخاطر اكتتاب مرتفعة، معدل الخسارة يساوي 1.20 فأكثر.

وبناء عليه تكون اللبيانات النهائية للمثال الاتوضيحي على الشكل التالي:

الشركة	X1	X2	X3	X4	Y
1	Exp3	RA	(ReC+)	LR2	(-Ret)
2	Exp3	RT	(ReC+)	LR2	(-Ret)
3	Exp3	RA	(ReC+)	LR1	(+Ret)
4	Exp1	RA	(ReC-)	LR3	(+Ret)
5	Exp1	RT	(ReC-)	LR1	(+Ret)
6	Exp2	RA	(ReC+)	LR2	(-Ret)
7	Exp1	RA	(ReC-)	LR2	(+Ret)
8	Exp2	RA	(ReC-)	LR3	(+Ret)
9	Exp2	RT	(ReC-)	LR2	(+Ret)
10	Exp2	RT	(ReC+)	LR1	(+Ret)
11	Exp3	RA	(ReC-)	LR1	(+Ret)
12	Exp2	RT	(ReC+)	LR3	(-Ret)
13	Exp1	RT	(ReC-)	LR3	(-Ret)
14	Exp2	RA	(ReC+)	LR3	(+Ret)

يتم حساب مقياس Entropy لكل متغير وفقا للعلاقة التالية:

الخطوة الأولى: يتم حساب قيمة $Entropy\ (Y)$ للمتغير التابع:

يلاحظ من الجدول السابق، بالنسبة للمتغير التابع (Y) أن عدد الشركات التي يبلغ نسبة الاحتفاظ بالأقساط لديها أكبر من ٤٠% (Ret+) هو (٩) شركات. بينما عدد الشركات التي يبلغ نسبة الاحتفاظ بالأقساط لديها أقل من أو يساوي ٤٠% (Ret−) هو (٥) شركات، أي أن:

$$(Y) = (9^{+}, 5^{-})$$

وبالتالي فإن:

$$Entropy\ (Y) = -\frac{9}{14}.Log_2\frac{9}{14} - \frac{5}{14}.Log_2\frac{5}{14} = 0.9402$$

الخطوة الثانية: اختيار المتغير المستقل الذي سوف يمثل جذر الشجرة Tree Root :

المتغير المستقل الذي سوف يمثل جذر الشجرة هو المتغير الذي يقدم أكبر تخفيض ممكن لعدم التأكد، وبالتالي الحصول على أكبر قدر من المعلومات التي سوف تساعدنا في التصنيف أو إتخاذ القرار، ويتم تحديد ذلك حساب مقياس Information Gain لكل متغير من المتغيرات المستقلة كما يلي:

أولا: يتم حساب $Entropy\ (Y, Xi)$ ، لكل متغير من المتغيرات المستقلة:

(١) بالنسبة للمتغير X2:

$$(X2_{RA}) = (6^{+}, 2^{-})$$

$$Entropy\ (X2_{RA}) = -\frac{6}{8}.Log_2\frac{6}{8} - \frac{2}{8}.Log_2\frac{2}{8} = 0.81127$$

$$(X2_{RT}) = (3^{+}, 3^{-})$$

$$Entropy\ (X2_{RT}) = -\frac{3}{6}.Log_2\frac{3}{6} - \frac{3}{6}.Log_2\frac{3}{6} = 0.9999$$

(٢) بالنسبة للمتغير X4:

$$(X4_{LR2}) = (2^{+}, 3^{-})$$

$$Entropy\ (X4_{LR2}) = -\frac{2}{5}.Log_2\frac{2}{5} - \frac{3}{5}.Log_2\frac{3}{5} = 0.9707$$

$$(X4_{LR1}) = (4^{+}, 0^{-})$$

$$Entropy\ (X4_{LR1}) = -\frac{4}{4}.Log_2\frac{4}{4} - \frac{0}{4}.Log_2\frac{0}{4} = 0$$

$$(X4_{LR3}) = (3^+, 2^-)$$

$$Entropy\ (X4_{LR3}) = -\frac{3}{5}.Log_2\frac{3}{5} - \frac{2}{5}.Log_2\frac{2}{5} = 0.9707$$

(٣) بالنسبة للمتغير X3:

$$(X3_{+ReC}) = (3^+, 4^-)$$

$$Entropy\ (X3_{+ReC}) = -\frac{3}{7}.Log_2\frac{3}{7} - \frac{4}{7}.Log_2\frac{4}{7} = 0.98522$$

$$(X3_{-ReC}) = (6^+, 1^-)$$

$$Entropy\ (X3_{-ReC}) = -\frac{6}{7}.Log_2\frac{6}{7} - \frac{1}{7}.Log_2\frac{1}{7} = 0.59167$$

(٤) بالنسبة للمتغير X1:

$$(X1_{Exp3}) = (2^+, 2^-)$$

$$Entropy\ (X1_{Exp3}) = -\frac{2}{4}.Log_2\frac{2}{4} - \frac{2}{4}.Log_2\frac{2}{4} = 0.9999$$

$$(X1_{Exp2}) = (4^+, 2^-)$$

$$Entropy\ (X1_{Exp2}) = -\frac{4}{6}.Log_2\frac{4}{6} - \frac{2}{6}.Log_2\frac{2}{6} = 0.91829$$

$$(X1_{Exp1}) = (3^+, 1^-)$$

$$Entropy\ (X1_{Exp1}) = -\frac{3}{4}.Log_2\frac{3}{4} - \frac{1}{4}.Log_2\frac{1}{4} = 0.811278$$

ثانيا: يتم حساب $Info\ Gain\ (S, X)$ ، لكل متغير من المتغيرات المستقلة، كما يلي:

$$Info\ Gain\ (Y, X) = Entropy\ (y) - \sum_{i=1}^{n}\frac{|S_v|}{S}.Entropy\ (S_v)$$

(١) بالنسبة للمتغير X2:

$$Info\ Gain\ (Y, X2)$$

$$= Entropy\ (Y) - \frac{8}{14}.Entropy\ (X2_{RA}) - \frac{6}{14}.Entropy\ (X2_{RT})$$

$$= 0.9402 - \frac{8}{14}.(0.81127) - \frac{6}{14}.(0.9999)$$

$$= 0.048$$

(٢) بالنسبة للمتغير X4:

$$Info\ Gain\ (Y, X4)$$
$$= Entropy\ (Y) - \frac{5}{14}.Entropy\ (X4_{LR2})$$
$$- \frac{4}{14}.Entropy\ (X4_{LR1}) - \frac{5}{14}.Entropy\ (X4_{LR3})$$
$$= 0.9402 - \frac{5}{14}.(0.9707) - \frac{4}{14}.(0) - \frac{5}{14}.(0.9707)$$
$$= 0.24686$$

(٣) بالنسبة للمتغير X3:

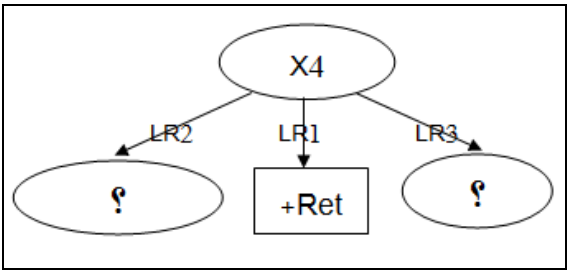
$$Info\ Gain\ (Y, X3)$$
$$= Entropy\ (Y) - \frac{7}{14}.Entropy\ (X3_{+ReC})$$
$$- \frac{7}{14}.Entropy\ (X3_{-ReC})$$
$$= 0.9402 - \frac{7}{14}.(0.9852) - \frac{7}{14}.(0.59167)$$
$$= 0.151765$$

(٤) بالنسبة للمتغير X1:

$$Info\ Gain\ (Y, X1)$$
$$= Entropy\ (Y) - \frac{4}{14}.Entropy\ (X1_{Exp3})$$
$$- \frac{6}{14}.Entropy\ (X1_{Exp2}) - \frac{4}{14}.Entropy\ (X1_{Exp1})$$
$$= 0.9402 - \frac{4}{14}.(0.9999) - \frac{6}{14}.(0.91829)$$
$$- \frac{4}{14}.(0.811278)$$
$$= 0.029256$$

يتضح مما سبق أن أكبر Info Gain كان للمتغير مخاطر الاكتتاب (X4)، وبالتالي فهو أفضل مصنف Best Classifier للمتغير التابع، وبناء عليه سيكون هو الممثل لجذر الشجرة Root Of The Tree. ويلاحظ بالنسبة للمتغير X4 أن الحالة الخاصة بـ LR1 كان

$Entropy\ (X4_{LR1})$ يساوي الصفر، هذا معناه أنها سوف تصبح Leaf Node لشجرة القرار، كما هو موضح بالشكل التالي:



الشكل رقم (٢)

شجرة القرارات بعد المرحلة الأولي

في حين أن الحالتين الباقيتين (LR2, LR3) مازلنا في حاجة إلى المزيد من التصنيف، لذا سيتم البحث عن أفضل متغير مصنف لكل حالة من هذه الحالة بالاعتماد على قيمة Info Gain لكل متغير، كما يلي:

أولا: بالنسبة للحالة LR2:

(1) المتغير X2:

$$Entropy(LR2,RA) = (1^+,2^-)$$
$$= -\frac{1}{3}.Log_2\frac{1}{3} - \frac{2}{3}.Log_2\frac{2}{3} = 0.918295$$
$$Entropy(LR2,RT) = (1^+,1^-)$$
$$= -\frac{1}{2}.Log_2\frac{1}{2} - \frac{1}{2}.Log_2\frac{1}{2} = 0.9999$$
$$Info\ Gain\ (LR2,X2) = 0.9707 - \frac{3}{5}(0.918295) - \frac{2}{5}(0.9999)$$
$$= 0.19728$$

(٢) المتغير X3:

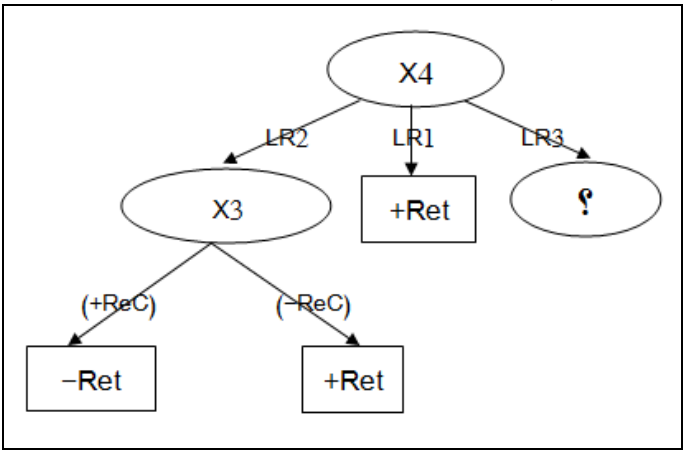
$$Entropy(LR2,+ReC) = (0^+,3^-)$$
$$= -\frac{0}{3}.Log_2\frac{0}{3} - \frac{3}{3}.Log_2\frac{3}{3} = 0$$
$$Entropy(LR2,-ReC) = (2^+,0^-)$$
$$= -\frac{2}{2}.Log_2\frac{2}{2} - \frac{0}{2}.Log_2\frac{0}{2} = 0$$

$$Info\ Gain\ (LR2, X3) = 0.9707 - \frac{3}{5}(0) - \frac{2}{5}(0)$$
$$= 0.9707$$

(3) المتغير X1:

$$Entropy(LR2, Exp3) = (0^+, 2^-)$$
$$= -\frac{0}{2}.Log_2\frac{0}{2} - \frac{2}{2}.Log_2\frac{2}{2} = 0$$
$$Entropy(LR2, Exp1) = (1^+, 0^-)$$
$$= -\frac{1}{2}.Log_2\frac{1}{2} - \frac{0}{2}.Log_2\frac{0}{2} = 0$$
$$Entropy(LR2, Exp2) = (1^+, 1^-)$$
$$= -\frac{1}{2}.Log_2\frac{1}{2} - \frac{1}{2}.Log_2\frac{1}{2} = 0.9999$$
$$Info\ Gain\ (LR2, X1) = 0.9707 - \frac{2}{5}(0) - \frac{1}{5}(0) - \frac{2}{5}(0.9999)$$
$$= 0.5707$$

يتضح مما سبق أن أكبر Info Gain كان للمتغير X3، وبالتالي فهو أفضل مصنف Best Classifier للحالة LR2. ويلاحظ بالنسبة للمتغير X3 أن حالتي هذا المتغير كانت قيمة الـ $Entropy$ تساوي الصفر، هذا معناه أنهما سوف يصبحان Leaf Nodes لشجرة القرار، كما هو موضح بالشكل التالي:



الشكل رقم (٣)

شجرة القرارات بعد المرحلة الثانية

ثانيا: بالنسبة للحالة LR3:

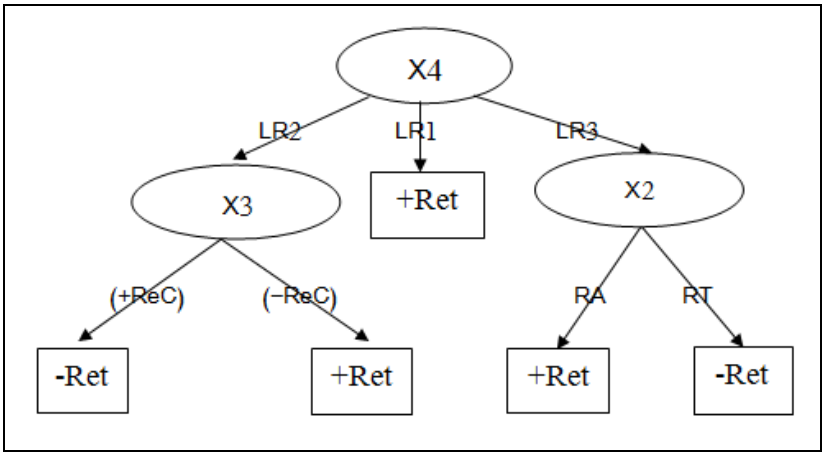
(1) المتغير X2:

$$Entropy(LR3, RA) = (3^+, 0^-)$$
$$= -\frac{3}{3}.Log_2\frac{3}{3} - \frac{0}{3}.Log_2\frac{0}{3} = 0$$
$$Entropy(LR3, RT) = (0^+, 2^-)$$
$$= -\frac{0}{2}.Log_2\frac{0}{2} - \frac{2}{2}.Log_2\frac{2}{2} = 0$$
$$Info\ Gain\ (LR3, X2) = 0.9707 - \frac{3}{5}(0) - \frac{2}{5}(0)$$
$$= 0.9707$$

(٢) المتغير X1:

$$Entropy(LR3, Exp3) = (0^+, 0^-)$$
$$= -\frac{0}{2}.Log_2\frac{0}{2} - \frac{0}{2}.Log_2\frac{0}{2} = 0$$
$$Entropy(LR3, Exp1) = (1^+, 1^-)$$
$$= -\frac{1}{2}.Log_2\frac{1}{2} - \frac{1}{2}.Log_2\frac{1}{2} = 0.9999$$
$$Entropy(LR3, Exp2) = (2^+, 1^-)$$
$$= -\frac{2}{2}.Log_2\frac{2}{2} - \frac{1}{2}.Log_2\frac{1}{2} = 0.91829$$
$$Info\ Gain\ (LR3, X1) = 0.9707 - \frac{2}{5}(0.9999) - \frac{3}{5}(0.91829)$$
$$= 0.0197$$

يلاحظ هنا أن أكبر Info Gain كان للمتغير X2 وبالتالي فهو أفضل مصنف Best Classifier للحالة LR3. ويلاحظ بالنسبة لهذا المتغير أن حالتي هذا المتغير كانت قيمة $Entropy$ تساوي الصفر، هذا معناه أنهما سوف يصبحان Leaf Nodes لشجرة القرار، وبذلك تكون شجرة القرارات النهائية على الشكل التالي:



الشكل رقم (٤)

شجرة القرارات بعد المرحلة الثالثة والنهائية

يتضح من الشكل السابق أن هناك ٥ مسارات لمستوى الاحتفاظ المتوقع:

(١) إذا كان معدل الخسارة في محفظة الاكتتابات الخاصة بشركة التأمين (LR) أقل من ٩٠%، فإن مستوى الاحتفاظ بالأقساط المتوقع سيكون أكبر من ٤٠%.

(٢) إذا كان معدل الخسارة في محفظة الاكتتابات الخاصة بشركة التأمين (LR) يتراوح بين ٩٠% إلى أقل من ١٢٠%، وكان تكلفة إتفاقيات إعادة التأمين أكبر من ١٠٠%، فإن مستوى الاحتفاظ بالأقساط المتوقع سيكون أقل من ٤٠%.

(٣) إذا كان معدل الخسارة في محفظة الاكتتابات الخاصة بشركة التأمين (LR) من ٩٠% إلى أقل من ١٢٠%، وكان تكلفة إتفاقيات إعادة التأمين أقل من أو تساوي ١٠٠%، فإن مستوى الاحتفاظ بالأقساط المتوقع سيكون أكبر من ٤٠%.

(٤) إذا كان معدل الخسارة في محفظة الاكتتابات الخاصة بشركة التأمين (LR) أكبر من أو يساوي ١٢٠%، وكان معدل العائد على الاستثمار منخفض، فإن مستوى الاحتفاظ بالأقساط المتوقع سيكون أقل من ٤٠%.

(٥) إذا كان معدل الخسارة في محفظة الاكتتابات الخاصة بشركة التأمين (LR) أكبر من أو يساوي ١٢٠%، وكان معدل العائد على الاستثمار مرتفع، فإن مستوى الاحتفاظ بالأقساط المتوقع سيكون أكبر من ٤٠%.

٧. الدراسة التطبيقية:

٧.١ البيانات ومصادر الحصول عليها:

بالنسبة للمتغيرات التأمينية، فقد تم الحصول على البيانات الخاصة بها من واقع البيانات المنشورة في الكتاب الإحصاء السنوى الصادر عن الهيئة المصرية للرقابة على التأمين خلال فترة الدراسة من ١٩٩٩/٢٠٠٠ حتى ٢٠١٦/٢٠١٧. أما البيانات الاقتصادية، بالنسبة للمتغير الخاص بسعر الصرف من البنك المركزي المصري، أما بالنسبة للمتغيرين معدل النمو الاقتصادي ومعدل التضخم فقد تم الحصول عليهما من شبكة المعلومات الدولية من الموقع التالي: https://tradingeconomics.com/egypt/ .

٧.٢ نتائج التحليل الإحصائي:

تم الاعتماد على برنامج WEKA للتنقيب عن البيانات عند بناء شجرة القرارات التصنيفية للتنبؤ بمستوى الاحتفاظ بالأقساط في سوق التأمين المصري باستخدام خوارزمية C4.5، ومما هو جدير بالذكر أن خوارزمية C4.5 في برنامج WEKA تسمى Java 48 أو J48 [٤٥]. وقد قام الباحث بتشغيل البيانات في حالتين: الحالة الأولي في حالة قبل استبعاد المتغيرات الاقتصادية العامة، والحالة الثانية بعد استبعاد المتغيرات الاقتصادية العامة والاكتفاء بالمتغيرات التأمينية فقط

وبصفة عامة، للحكم على مدى صلاحية النموذج للتطبيق سوف يتم الاعتماد على أكثر من مقياس من مقاييس دقة التصنيف سواء على مستوى النموذج ككل أو على مستوى كل فئة من فئات المخرجات، كما يلي:

(١) نسبة التصنيف الصحيح للحالات أو السجلات محل الدراسة: ويلاحظ من نتائج التحليل الاحصائي الموضحة بالجدول رقم (....)، أن نسبة نجاح خوارزمية C4.5 في تصنيف الفئات الخاصة بالمخرجات (مستوى الاحتفاظ بالأقساط) بشكل صحيح كانت ٧٧.٨%، حيث أنها استطاعت أن تصنف عدد ١٤ حالة أو سجل من السجلات محل الدراسة بشكل صحيح، في حين أنها فشلت في تصنيف ٤ حالات. وهذه النسبة تعتبر نسبة جيدة ومقبولة ودليل على الدقة العالية للنموذج في عملية التنبؤ بمستوى الاحتفاظ بالأقساط في سوق التأمين المصري.

[٤٥] Chauhan, Harvinder and Chauhan, Anu, Op. cit.

(٢) معامل Kappa : يلاحظ أن قيمة هذه المعامل كما هو موضح بالجدول رقم (٣) بالملاحق هي ٦٤.٢% وهي أيضا نسبة مقبولة ودليل على الدقة العالية للنموذج في التنبؤ، حيث تتراوح قيمة هذه المعامل (-١ ، +١).

الجدول رقم (٣)

مقاييس دقة التنبؤ الكلية لشجرة القرارات في سوق التأمين المصري

عدد السجلات المصنفة بشكل صحيح	%	عدد السجلات المصنفة بشكل غير صحيح	%	العدد الكلي للسجلات	%	إحصائي *Kappa*
١٤	٧٧.٨	٤	١١.١	١٨	١٠٠	٠.٦٤٢

(٣) مصفوفة التشويش Confusion Matrix : توضح هذه المصفوفة طريقة تصنيف السجلات وتوزيعها بين الحالات المختلفة أو الفئات المختلفة للمخرجات (مستويات الاحتفاظ بالأقساط). كما هو موضح بالشكل التالي:

```
=== Confusion Matrix ===

 a b c   <-- classified as
 4 0 1 | a = a
 1 2 2 | b = b
 0 0 8 | c = c
```

الشكل رقم (٥) مصفوفة التشويش

يتم الاعتماد على هذه المصفوفة في حساب ٤ معدلات أو معاملات أساسية في عملية التقييم – كما هو موضح بالشكل رقم (...) – يتم الاعتماد عليها في حساب جميع مقاييس دقة التنبؤ أو التصنيف وفقا لأسلوب شجرة البيانات سواء بشكل مباشر أو غير مباشر، وهذه المعدلات هي:

(أ) عدد التصنيف الموجب الصحيح True Positive (TP): هو عبارة عن عدد السجلات التي تنتمي الى فئة معينة من فئات المخرجات (المتغير التابع) وتم تصنيفها في نفس الفئة.

(ب) عدد التصنيف السالب الخاطئ False Negative (FN) : هو عبارة عن عدد السجلات التي تنتمي الى فئة معينة، ولكن تم تصنيفها في فئات أخرى .

(ج) عدد التصنيف الموجب الخاطئ False Positive (FP): ويعرف بالخطأ من النوع الأول Type | Error، وهو عبارة عن عدد السجلات التي لا تنتمي الى فئة معينة من فئات المخرجات (المتغير التابع) وتم تصنيفها للفئة.

(د) عدد التصنيف السالب الصحيح True Negative (TN): ويعرف بالخطأ من النوع الثاني Type || Error، وهو عبارة عن عدد السجلات التي لا تنتمي الى فئة معينة من فئات المخرجات (المتغير التابع) وتم تصنيفها للفئات الأخرى.

وبحساب القيم السابقة لكل فئة من المخرجات نجد أنه:

بالنسبة للمستوى الأول للاحتفاظ (A):

True Positive (TP)=4 False Positive (FP)=1

False Negative (FN)=1 True Negative (TN)=12

بالنسبة للمستوى الثاني للاحتفاظ (B):

True Positive (TP)=2 False Positive (FP)=0

False Negative (FN)=3 True Negative (TN)=13

بالنسبة للمستوى الثالث للاحتفاظ (C):

True Positive (TP)=8 False Positive (FP)=3

False Negative (FN)=0 True Negative (TN)=7

وبناء على القيم السابقة يتم حساب عدد من معاملات أو مقاييس دقة التصنيف التي يتم الاعتماد عليها في شجرة القرارات التصنيفية الموضحة بالجدول التالي:

الجدول رقم (٤)

بعض مقاييس دقة التصنيف في شجرة القرارات التصنيفية

Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
A	0.80	0.077	0.800	0.800	0.862
B	0.40	0.000	1.000	0.571	0.538
C	1.00	0.300	0.727	0.842	0.719
المتوسط المرجح	0.778	0.155	0.823	0.755	0.708

يتضح من الجدول السابق:

(أ) معامل التصنيف الصحيح True Positive Rate (TPR): هو عبارة عن نسبة السجلات التي تنتمي الى فئة معينة من فئات المخرجات (المتغير التابع) وتم تصنيفها في نفس الفئة. ويتم حساب هذا المعامل كما يلي:

$$TP\ Rate = \frac{TP}{TP + FN}$$

$$TP\ Rate\ (A) = \frac{4}{4+1} = \frac{4}{5} = 0.80$$

$$TP\ Rate\ (B) = \frac{2}{2+3} = \frac{2}{5} = 0.40$$

$$TP\ Rate\ (C) = \frac{8}{8+0} = \frac{8}{8} = 1.00$$

وكان المعدل العام أو المتوسط ٧٧.٨% وهي نسبة مقبولة لعملية التصنيف.

(ب) معامل التصنيف الخاطئ False Positive Rate (FPR): هو عبارة عن عن نسبة السجلات التي لا تنتمي الى فئة معينة من فئات المخرجات (المتغير التابع) وتم تصنيفها للفئة. ويتم حساب هذا المعامل كما يلي:

$$FP\ Rate = \frac{FP}{FP + TN}$$

$$FP\ Rate\ (A) = \frac{1}{1+12} = \frac{1}{13} = 0.076923 = 0.077$$

$$FP\ Rate\ (B) = \frac{0}{0+13} = \frac{0}{13} = 0$$

$$FP\ Rate\ (C) = \frac{3}{3+7} = \frac{3}{10} = 0.30$$

وكان المعدل العام أو المتوسط ٨٢.٣% وهي نسبة جيدة ومقبولة لعملية التصنيف.

(ج) معامل التدقيق Precision : يتم حساب هذا المعامل أو المقياس من خلال العلاقة التالية:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision\ (A) = \frac{4}{4+1} = 0.80$$

$$Precision\ (B) = \frac{2}{2+0} = 1.00$$

$$Precision\ (C) = \frac{8}{8+3} = 0.727$$

وكان المعدل العام أو المتوسط ٧٧.٨% وهي نسبة جيدة للاطمئنان لعملية التصنيف.

(د) مقياس ف F–Measure : هو مقياس إجمالي لدقة الخوارزمية المستخدمة في عملية التصنيف، ويلخص كل من Precision و Recall، علما بأن معامل الاستدعاء Recall هي نفسها TPR، يتم حساب هذا المعامل أو المقياس بإحدى الطريقتين التاليتين كما يلي:

$$F-Measure=\frac{2TP}{2TP+FP+FN}$$

or

$$F-Measure=\frac{2}{\frac{1}{Recall}+\frac{1}{Precision}}$$

$$F-Measure\ (A)=\frac{2}{\frac{1}{0.80}+\frac{1}{0.80}}=0.80$$

$$F-Measure\ (B)=\frac{2}{\frac{1}{0.40}+1}=0.5714$$

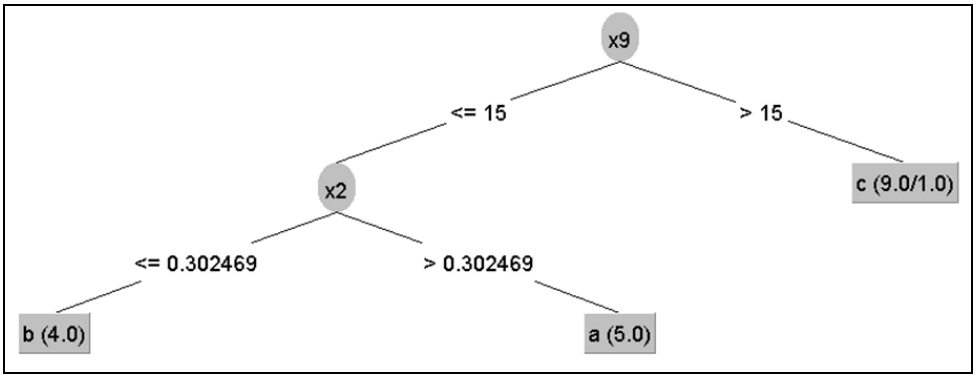
$$F-Measure\ (C)=\frac{2}{1+\frac{1}{0.727}}=0.8419$$

وكان المعدل العام أو المتوسط لهذا المقياس ٧٥.٥% وهي أيضا نسبة جيدة ومقبولة لعملية التصنيف والاطمئنان لعلملية التنبؤ.

(هـ) معامل ROC Area : تعبر عن المساحة أسفل المنحنى المرسوم بين FPR على المحور الأفقي، و TPR على المحور الرأسي، وغني عن البيان أن قيمة هذا المعامل تتراوح بين الصفر والواحد الصحيح، وكلما كانت قيمته أكبر من ٠.٥ دليل على دقة التصنيف وأن التنبؤ من خلال النموذج ليس تنبؤ عشوائي Random Predication [٤٦]، ويلاحظ هنا أن قيمة هذا المعامل للنموذج ككل ٠.٧٠٨ وهي نسبة مقبولة ودليل على دقة التنبؤات التي يتم الحصول عليها من هذا النموذج والخاصة بتصنيف مستويات الاحتفاظ في سوق التأمين المصري.

والشكل رقم (٦) يوضح شجرة القرارات التصنيفية وفقا لخوارزمية C4.5 للتنبؤ بمستوى الاحتفاظ بالأقساط في سوق التأمين المصري، كما يلي:

[46] Fawcett, Tom (2004). "ROC Graphs: Notes and Practical Considerations for Researchers", Pattern Recognition Letters. 27 (8): 882–891.



الشكل رقم (٦)

شجرة القرارات التصنيفية للتنبؤ بالاحتفاظ بالأقساط

لسوق تأمينات الممتلكات والمسؤولية في مصر

يتضح من الشكل السابق أن هناك ٣ مسارات لمستوى الاحتفاظ المتوقع في سوق تأمينات الممتلكات والمسؤولية في سوق التأمين المصري، كما يلي:

(١) إذا كان عدد شركات التأمين العاملة في سوق التأمين المصري أكبر من ١٥ شركة سيكون مستوى الاحتفاظ بالأقساط سيكون كبير (أي نسب الاحتفاظ ستكون أكبر من ٥٠%)

(٢) أما لو كان عدد الشركات أقل من ١٥ شركة وكانت الرافعة المالية في سوق التأمين المصري أقل من أو يساوي ٠.٣٠ فإن نسب الاحتفاظ بالأقساط سوف تتراوح بين ٤٥% الى أقل من ٥٠%.

(٣) أما لو كان عدد الشركات أقل من ١٥ شركة وكانت الرافعة المالية في سوق التأمين المصري أكبر من ٠.٣٠ ، فإن نسب الاحتفاظ بالأقساط سوف تتراوح بين ٤٥% الى أقل من ٥٠%.

٨. النتائج والتوصيات :

٨.١ النتائج:

(أ) زيادة عدد شركات التأمين في سوق التأمين المصري يساعد على زيادة الاحتفاظ بالأقساط.

(ب) انخفاض الرافعة المالية يساعد أيضا على زيادة مستوى الاحتفاظ بالأقساط في سوق التأمين المصري.

٨.٢ التوصيات:

يوصي الباحث بزيادة الاعتماد من جانب متخذي القرار في شركات التأمين وهيئات الرقابة على التأمين بزيادة الاعتماد على أساليب التنقيب عن البيانات بصفة عامة وشجرة القرارات التصنيفية

على وجه الخصوص في التنبؤ بمستوى الاحتفاظ بالأقساط، لما تتميز به من سهولة في الفهم والتطبيق خاصة وأنها يتم تطبيقها في قطاع البنوك.

المراجع:

أولا المراجع العربية:

١. القاضي، عبد الحليم عبد الله، وآخرون، دراسة الطاقة الاستيعابية لسوق التأمين، أكاديمية البحث العلمي، ٢٠٠٤.

٢. عبد البارئ، محمد وحيد، حسن، نادية احمد، استخدام الاساليب الكمية لتحديد جد الاحتفاظ فى تأمينات الممتلكات والمسئولية المدنية تجاه الغير بالتطبيق على فرع الحريق المجلة المصرية للدراسات التجارية كلية التجارة– جامعة المنصورة – العدد الثانى – الجزء الثانى، ١٩٩٩.

٣. النادي، هويدا جمال الدين، دراسة تحليلية للعوامل المؤثرة في تحديد حدود الاحتفاظ في التأمينات العامة (بالتطبيق على فرع النقل البحري)، رسالة ماجستير غير منشورة، كلية التجارة، جامعة القاهرة، ٢٠٠٤.

٤. أحمد، مصطفى عبد الغني، الاتجاهات الحديثة لقياس الطاقة الاستيعابية ومعدل الاحتفاظ، المجلة المصرية للدراسات التجارية، مج ٢٤، ع١، ٢٠٠٠.

٥. عفيفي، زينب حسن، العوامل المؤثرة في الطاقة الاحتفاظية الصافية بالتأمينات العامة، دراسة عاملية تطبيقية على شركات التأمين المباشر، مجلة الدراسات المالية والتجارية، كلية التجارة – جامعة أسيوط، ١٩٨٨.

٦. عبد الجواد، عمر أحمد، إعداد نماذج حدود الاحتفاظ الأمثل باستخدام الأساليب الكمية – دراسة تحليلية بالتطبيق على تأمينات الحياة في السوق المصري، رسالة دكتوراه غير منشورة، كلية التجارة – جامعة المنصورة، ١٩٩٩.

٧. شهاب الدين، محمد مصطفى – مشعال، محمود عبد العال، استخدام أسلوب البوتسترابي في تقدير النموذج الأمثل لدالة الاحتفاظ بالأقساط في سوق التأمين السعودي، مجلة البحوث المعاصرة، كلية التجارة – جامعة سوهاج، ٢٠١٣

٨. عبد المولى، محمد – المهدي، محمد، نحو إطار متكامل لتحديد العوامل المؤثرة على حدود الاحتفاظ في التأمين على الحياة، دراسة تطبيقية على شركات التأمين المباشر العربية، مجلة آفاق جديدة، كلية التجارة – جامعة المنوفية،العدد الثاني، ١٩٩٥.

٩. إبراهيم، أحمد عبد الرحمن أحمد، استخدام نموذج تحليل المسار في تحديد أهم العوامل المؤثرة على معدل الاحتفاظ لشركات التأمين المصرية –دراسة تطبيقية، مجلة البحوث المالية والتجارية، كلية التجارة – جامعة بورسعيد، العدد الثاني، ٢٠١٢.

ثانيا: أولا المراجع الأجنبية:

10. Adams, M., Buckle, M., (2003),The determinants of corporate financial performance in the Bermuda insurance market. Appl. Finance. Econ., 13: 133-143.
11. Black K, Skipper HD (1994), Life Insurance. 12th ed., Prentice Hall, New York.
12. Browne MJ, Hoyt RE (1995),Economics and Market Predictors and Insolvencies in the Property-Liability insurance Industry. J. Risk Insur., 62(2): 309-327.
13. Carter RL (1979), Reinsurance. London: Kluwer Publishing.
14. Chauhan, Harvinder and Chauhan, Anu, Implementation of decision tree algorithm c4.5, International Journal of Scientific and Research Publications, Volume 3, Issue 10, October 2013.
15. Chen R, Wong KA (2004),The Determinants of Financial health of Asian Insurance Companies. J. Risk Insur., 71(3): 469-499.
16. Cummins JD, Dionne G Gagné R, Nouira A, Song QF (2008), The Costs and Benefits of Reinsurance. Working Paper, SCOR/JRI conference.
17. Cummins JD, Nini G P (2002), Optimal Capital Utilization by Financial Firms: Evidence from the Property-Liability Insurance Industry. J. Financ. Serv. Res., 21(1): 15-53.
18. Cummins, David, Sommer, David, (1996), Capital and risk in property-liability insurance markets. Journal of Banking & Finance, vol. 20, issue 6, 1069-1092.

19. *Dang, Chongyu , and Li, Frank , (2015), Measuring Firm Size in Empirical Corporate Finance, Journal of Banking & Finance Volume 86, January 2018, Pages 159-176.*
20. Fawcett, Tom (2004). "ROC Graphs: Notes and Practical Considerations for Researchers", Pattern Recognition Letters. 27 (8): 882–891.
21. Gatzlaff K (2009),Dimensions of Property- Liability insurer Performance. The Florida State University doctoral Dissertation.
22. Gottheimer, George M ,(1983), A Strategy for Reinsurance Management, CPCU J., 36: 206-218.
23. Grace, Martin F. and Hotchkiss, Julie L., External Impacts on the Property-Liability Insurance Cycle, Journal of Risk and Insurance 62 (December 1995): 738-54., 2011
24. Guo, Ligia, Applying Data Mining Techniques in Property and Causality Insurance, https://www.casact.org/pubs/forum/03wforum/03wf001.pdf.
25. Hammond, J. D., Shapiro, Arnold F.,(1979), Capital Requirements for Entry into Property and Liability Underwriting: An Empirical Examination, , APR 75-16550 A01, P.38.
26. Hssina, Badr and *et al*, A comparative study of decision tree ID3 and C4.5, International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications. www.ijacsa.thesai.org
27. Jang, Jonghag (2001), Reinsurance retentions and limits for property- Liability Insurers: Theory and Empirical Tests. Temple University doctoral Dissertation.
28. Lee K, Palmer BA, Skipper HD (1992), An Analysis of Life Insurer Retention imits. Journal of Risk and Insurance, 59(1):57-71.
29. Lee, Hsu-Hua and Lee, Chen-Ying, Determinants of property-liability insurer retention: Evidence from Taiwan insurance industry, African Journal of Business Management Vol. 5(32), pp. 12543-12550,14 December, 2011
30. Lee, Kwangbong and et al, (1992), An analysis of life insurer retention limits, Journal of Risk and Insurance, March, Volume: 59, Issue : 1.
31. Lior Rokach and Oded Maimon, Data Mining with Decision Trees: Theory and Applications, World Scientific Publishing, 2008.

32. Max Bramer, Principles of Data Mining, 3rd edition, Springer, Nov 9, 2016.
33. Mayers D, Smith C., (1990), On the Corporate Demand for Insurance: Evidence from Reinsurance Market. J. Bus., 63: 19-40.
34. Qin, Zengchang, and Tang, Yongchuan, (2014), Uncertainty Modeling for Data Mining: A Label Semantics Approach, Springer, P.78.
35. Quinlan, J. R. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (Mar. 1986), 81–106.
36. Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
37. Shannon, Claude E. (July–October 1948), "A Mathematical Theory of Communication", Bell System Technical Journal, 27 (3): 379–423. http://cm.belllabs.com/cm/ms/what/shannday/paper.html
38. Sharma, Seema and et al, Classification Through Machine Learning Technique: C4.5 Algorithm based on Various Entropies, International Journal of Computer Applications, Volume 82 – No 16, November 2013.
39. Sommer D. (1996), The Impact of Firm Risk on Property-liability Insurance Price. J. Risk Insur., 63(3):501- 514.
40. Szczepanski, Chester J., Cooper, Craig A., An Analysis of Variations in Leverage Ratios Among Insurers, https://www.casact.org/pubs/dpp/dpp92/92dpp269.pdf.
41. Zengchang Qin, Yongchuan Tang, Uncertainty Modeling for Data Mining: A Label Semantics Approach, Springer, 2014.

الملاحق:

A less than 0.45 , B : 0.45-0.50, C: greater than

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     retention
Instances:    18
Attributes:   13
              x1
              x2
              x3
              x4
              x5
              x6
              x7
              x8
              x9
              x10
              x11
              x12
              y
Test mode:    10-fold cross-validation
```

0.50

```
J48 pruned tree
------------------

x9 <= 15
|   x2 <= 0.302469: b (4.0)
|   x2 > 0.302469: a (5.0)
x9 > 15: c (9.0/1.0)

Number of Leaves  :     3

Size of the tree :      5


Time taken to build model: 0.04 seconds
```

```
=== Summary ===

Correctly Classified Instances          14               77.7778 %
Incorrectly Classified Instances         4               22.2222 %
Kappa statistic                          0.6418
Mean absolute error                      0.1806
Root mean squared error                  0.3901
Relative absolute error                 40.483  %
Root relative squared error             81.5085 %
Total Number of Instances               18

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure   MCC      ROC Area  PRC Area  Class
                 0.800    0.077      0.800     0.800     0.800     0.723     0.862     0.696     a
                 0.400    0.000      1.000     0.400     0.571     0.570     0.538     0.567     b
                 1.000    0.300      0.727     1.000     0.842     0.714     0.719     0.668     c
Weighted Avg.    0.778    0.155      0.823     0.778     0.755     0.676     0.708     0.647

=== Confusion Matrix ===

 a b c   <-- classified as
 4 0 1 | a = a
 1 2 2 | b = b
 0 0 8 | c = c
```