النمذجة الإكتوارية للمطالبات في التأمينات العامة بإستخدام تقنيات التنقيب في البيانات (Data Mining) دراسة تطبيقية على فرع تأمين البترول بالسوق المصري

د. محمد محمود هاشم
مدرس بأكاديمية السادات للعلوم الادارية

ملخص البحث

يهدف البحث إلى التعرف على طبيعة البيانات ومقدارها وكيفية التعامل مع الأحجام الكبيرة منها تمهيداً لمعالجتها, والعمل على تحسين مستوى استخدامها لرفع الأداء الفعلي, وتطبيق كل من (خوارزمية C4.5و خوارزمية CART وخوارزمية Logistic Regression وخوارزمية NaÏve Bayes) في نمذجة المطالبات لفرع تأمين البترول, وعرض مجموعة من المقترحات لاستخدام الخوارزميات الأربعة السابقة في التنقيب في البيانات للوصول إلى النمذجة الاكتوارية للمطالبات في فرع تأمين البترول من خلال مقارنة نتائج التطبيق, وقد توصل البحث إلى إمكانية تطبيق الخوارزميات الأربعة المقترحة في نمذجة مطالبات تأمين البترول في **شركتى مصر للتأمين وقناة السويس للتأمين،** وهذا يؤكد صحة الفرض البحثي الأول والقائل بأنه يمكن "**استخدام خوارزميات مختلفة للتنقيب في البيانات في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول**"، وأن هناك اختلاف في نتائج تطبيق كل خوارزمية من الخوارزميات الأربعة مما يدل على عدم صحة الفرض البحثي الثاني والقائل بأنه "**لا يوجد اختلاف بين تطبيق النماذج المختلفة للخوارزميات المستخدمة في البحث في النمذجة الاكتوارية للمطالبات.**

الكلمات الافتتاحية

النمذجة الاكتوارية, المطالبات, التأمينات العامة, تأمين البترول, التنقيب في البيانات, شجرة القرارات, خوارزمية C4.5 خوارزمية CART , خوارزمية Logistic Regression ,خوارزمية NaÏve Bayes.

Actuarial Modeling of Claims in Non-Life Insurances Using Data Mining Techniques: An Applied Study on Insuring Petroleum in the Egyptian Market

Abstract

The aim of this research is to identify the nature and amount of data to facilitate dealing with large amounts of data for processing, and to improve the ability of insurance companies' management to use it for developing actual performance, and applying the following algorithms (C4.5, CART, Logistic Regression and NaÏve Bayes) in modeling claims in Petroleum Insurance, The possibility of applying the four algorithms to the petroleum insurance Companies has been reached, and this confirms the validity of the first hypothesis which says that different algorithms can be used for data mining in the actuarial modeling of claims in public insurances petroleum insurance field, also there is a difference in the results of applying each of the four algorithms, which indicated the invalidity of the second hypothesis which says that there is no difference in applying the different algorithms -mentioned before- in the actuarial modeling of claims in public insurances.

Keywords: Actuarial Modeling, Claims, Public Insurances, Petroleum Insurance, Data Mining, Decision Tree, C4.5 algorithm, CART algorithm, Logistic Regression algorithm, NaÏve Bayes algorithm.

١. الإطار النظري للبحث

١/١ المقدمة وطبيعة المشكلة

مع التطور الكبير في منظمات الأعمال وخصوصاً شركات التأمين، والزيادة الهائلة في كمية البيانات والمعلومات الناتجة عن هذا التطور،أصبحت الحاجة إلى إمتلاك تقنيات للتعامل مع قواعد البيانات الضخمة ضرورة ملحة, ولكن لم يلق هذا الموضوع إهتمام كافٍ من قبل إدارة معظم شركات التأمين والتوجه نحو إمتلاك التقنيات الحديثة التي تساعد على فهم قواعد البيانات وكيفية التعامل معها, ومن ثم التنقيب في بياناتها وصولاً للمعرفة المطلوبة وعلى النحو المؤثر في أداء تلك الشركات وقراراتها المتخذة, وجدير بالذكر أن استخدام تقنيات التنقيب في البيانات يوفر للمؤسسات في جميع المجالات القدرة على استكشاف أهم المعلومات في قواعد البيانات والتركيز عليها،وفي الوقت نفسه تساعد على بناء التنبؤات المستقبلية و اكتشاف السلوك و الاتجاهات مما يسمح باتخاذ القرارات الصحيحة واتخاذها في الوقت المناسب, ولقد تم استخدام تقنيات التنقيب في البيانات في مجال التأمين بصورة كبيرة وخاصة في توزيع العملاء بهدف تحديد أسعار وثائق التأمين في الفروع المختلفة, وتحديد التذبذب في المطالبات المتوقعة مستقبلاً أو نمذجة المطالبات، وأيضاً تحديد المطالبات الاحتيالية.

وفي ضوء المقدمة السابقة وفي ضوء سعي المنظمات والشركات المختلفة للبحث في مجال التنقيب في البيانات لإنتاج أدوات تساعد في التنقيب في البيانات، يمكن بلورة مشكلة البحث في التساؤلات الآتية:

- مامدى اهتمام إدارة شركات التأمين في تدوين وتوثيق بياناتها؟
- هل تعتمد إدارة شركات التأمين على الأساليب العلمية الحديثة في التعامل مع تلك البيانات؟
- ما مدى مساهمة كل من خوارزميات C4.5 ، CART ، Logistic Regression ، NaÏveBayes في النمذجة الاكتوارية للمطالبات في التأمينات العامة ؟

٢/١ أهداف البحث

يسعى البحث إلى تحقيق مجموعة من الأهداف أهمها:

١/٢/١ التعرف على طبيعة البيانات ومقدارها وكيفية التعامل مع الأحجام الكبيرة منها تمهيداً لمعالجتها.

٢/٢/١ العمل على تحسين مستوى استخدام قواعد البيانات من قِبل إدارة شركات التأمين لرفع الأداء الفعلي.

٣/٢/١ تطبيق خوارزمية C4.5 في نمذجة المطالبات في فرع تأمين البترول.

٤/٢/١ تطبيق خوارزمية CART في نمذجة المطالبات في فرع تأمين البترول.

٥/٢/١ تطبيق خوارزمية Logistic Regression في نمذجة المطالبات في فرع تأمين البترول.

٦/٢/١ تطبيق خوارزمية NaÏve Bayes في نمذجة المطالبات في فرع تأمين البترول.

٧/٢/١ عرض مجموعة من المقترحات لاستخدام الخوارزميات الأربعة السابقة في التنقيب في البيانات؛ للوصول إلى النمذجة الاكتوارية للمطالبات في فرع تأمين البترول من خلال مقارنة نتائج التطبيق.

٣/١ أهمية البحث

تهتم شركات التأمين بقرار تحديد قسط التأمين، وفي الوقت نفسه بنمذجة المطالبات والتنبؤ بها, وعند قيامها بالنمذجة الاكتوارية للمطالبات في فرع هام من فروع التأمين وهو تأمين البترول؛ فإنها من الممكن استخدام تقنيات حديثة مثل تقنية التنقيب في البيانات من خلال العديد من التطبيقات في هذا المجال, ولذا فان هذا البحث يكتسب أهميته من خلال ما يلي:

١/٣/١ أهمية البيانات باعتبارها المحرك الأساسي لعمل المنظمات خصوصاً شركات التأمين، والذي على أساسه تتحدد قواعد البيانات وبموجبه تعمل تقنية تنقيب البيانات.

٢/٣/١ أهمية قطاع التأمين باعتباره أحد أعمدة الاقتصاد القومي، والذي يلعب دوراً أساسياً في إجراء التحولات الاقتصادية على المستوى القومي.

٣/٣/١ أهمية إنشاء قاعدة بيانات خاصة بقطاع التأمين لتسهيل عملية توثيق البيانات وتنظيمها؛ لتصبح متسقة مع احتياجات حاملي الوثائق وإدارة الشركات على حدٍ سواء, ودعم قراراتهم المتخذة.

٤/٣/١ أهمية الخوارزميات الأربعة المستخدمة في البحث كونها توفر معالجة مستمرة للبيانات، وتتمتع بالمرونة، وتساهم في النمذجة الاكتوارية للمطالبات في تأمينات الممتلكات.

٥/٣/١ ضروة استفادة شركات التأمين المصرية من البيانات الضخمة المتاحة واستخدام النماذج الاكتوارية في تحديد سلوك المطالبات تمهيداً لتحديد السعر المناسب.

٤/١ فروض البحث

في ضوء مشكلة البحث وأهدافه قام الباحث بصياغة الفروض كما يلي::

١/٤/١ يمكن استخدام خوارزميات مختلفة للتنقيب في البيانات في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول.

٢/٤/١ لا يوجد اختلاف بين تطبيق النماذج المختلفة للخوارزميات المستخدمة في البحث في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول.

٥/١ منهجية البحث

تحقيقاً لهدف البحث قام الباحث باتباع عدة خطوات منهجية:

١/٥/١ المنهج الاستقرائي: ومن خلاله تم الاعتماد على المراجعة النظرية والفكرية لما كتب من دارسات وبحوث عربية وأجنبية, وذلك من خلال استعراض الدراسات السابقة سواء المرتبطة بموضوع المطالبات في التأمينات العامة أو تلك المرتبطة باستخدام تقنية التنقيب في البيانات.

٢/٥/١ المنهج التطبيقي: وتم ذلك من خلال استخدام تقنية التنقيب في البيانات في نمذجة مطالبات تأمين البترول في شركات التأمين المصرية,وقد اعتمد الباحث في هذا الجزء على تحليل البيانات المستخرجة من التقارير المنشورة خلال الفترة من ٢٠١١م حتى ٢٠٢٠م،وذلك باستخدام أربع خوارزميات مختلفة هي (خوارزمياتC4.5 وCART وخوارزمية Logistic Regression وخوارزمية NaÏve Bayes) في النمذجة الاكتوارية للمطالبات في التأمينات العامة.

٦/١ حدود البحث

١/٦/١ الحدود الزمنية: اقتصر البحث على الفترة من ٢٠١١ حتى ٢٠٢٠م.

٢/٦/١ الحدود المكانية: اقتصر البحث على شركتين فقط من شركات التأمين المصرية التي تعمل في فرع التأمين على البترول وهما: ***شركة مصر للتأمين وشركة قناة السويس للتأمين***، وذلك لتوافر البيانات الخاصة بهما وباعتبارهما يستحوذان على النسبة الأكبر من حصة السوق المصري من بين ٦ شركات تزاول هذا النوع من التأمين.

٣/٦/١ الحدود الموضوعية: اقتصر البحث على فرع تأمين البترول كونه من الفروع التي تتفاوت فيها المطالبات سواء من ناحية تكرارها أوقيمتها.

٢. الدراسات السابقة

في هذا الجزء من البحث يعرض الباحث أهم الدراسات السابقة سواء العربية أو الأجنبية التي تناولت موضوعات تتعلق بالمطالبات في فرع تأمين البترول، أو تلك المتعلقة باستخدام تقنية التنقيب في البيانات ثم التعليق عليها وتحليلها؛ بغرض الاستفادة مما توصلت إليه من نتائج وتوصيات تتعلق بموضوع البحث الحالي, والوصول إلى الفجوة البحثية والتي يتم بناء البحث عليها.

١/٢ الدراسات العربية

١/١/٢ دراسة (عطا, محمد محمد ,٢٠١٠)

تناولت الدراسة تقييم النشاط التأميني المباشر لقطاع البترول المصري والتنبؤ به بالتطبيق على خطر الحريق والانفجار , وقد تم تقييم النشاط التأميني لهذا القطاع من خلال سجلات شركات التأمين بهدف التوصل إلى نتائج تفيد شركات التأمين في إعادة النظر سواء بالنسبة لأسعار التأمين المطبقة على هذا القطاع أو بالنسبة لسياستها التسويقية, وقد هدفت الدراسة إلى دراسة وتحليل البيانات الخاصة بالنشاط التأميني لقطاع البترول المصري خلال الفترة من ١٩٩٦ إلى ٢٠٠٦م, مع بيان الوفورات المالية التي يمكن أن تحققها شركات التأمين والخاصة بقطاع البترول, وحث شركات التأمين المصرية على إعادة النظر في تقييم الأسعار بما يمكنها للوصول إلى السعر الذي يحقق لها المنافسة في ظل النظام الاقتصاد العالمي الجديد, وأخيراً التنبؤ بالنشاط التأميني لقطاع البترول في الفترة من ٢٠٠٦ إلى ٢٠١٢م، وبيان

الوفورات المتوقعة الخاصة به. وقد توصلت الدراسة إلى أن هناك ارتفاع في الأقساط المكتسبة لقطاع البترول خلال فترة الدراسة سببه نمو قطاع البترول وزيادة الاكتشافات البترولية الجديدة مع زيادة الوعي التأميني لدى المسئولين عن قطاع البترول, كما توصلت الدراسة إلى تزايد الرقم القياسي لإجمالي الأقساط. وأوصت الدراسة بضرورة إعادة النظر من جانب شركات التأمين في الأسعار الخاصة بتأمين الحريق والإنفجار بالنسبة لقطاع البترول, مع ضرورة منح بعض الشركات خصم عدم المطالبة, وحث بعض الشركات محل الدراسة وتشجيعها على التأمين ضد خطر الحريق والانفجار.

٢/١/٢ (دراسة سيد,أحمد فايز أحمد, ٢٠١٦)

ركزت هذه الدراسة على طرق التنقيب عن البيانات مفتوحة المصدر, وأوضحت أنه قد ظهر مع التنقيب في الكم الهائل والمتزايد من البيانات –للوصول إلى المعرفة في قواعد البيانات المختلفة– أدوات وبرمجيات تساعد على ذلك. وقد هدفت الدراسة إلى دراسة ماهية التنقيب عن البيانات ومهامه وتطبيقاته, وتحليل أدوات التنقيب عن البيانات مفتوحة المصدر وتقييمها, وأيضاً إجراء مقارنة بين التنقيب عن البيانات مفتوحة المصدر. وتوصلت الدراسة إلى عدة نتائج أهمها: وجود مزايا تتصف بها بعض الأدوات في التنقيب في البيانات عن غيرها, وأن هناك أربع أدوات متوفرة للتنقيب في البيانات هي:RapidMiner,Weka, Orange,KNIMA.

٣/١/٢ دراسة (البكري, محمد عبد السلام عبد العزيز, ٢٠١٦)

قدمت هذه الدراسة أحد الأساليب الهامة والتي انتشر تطبيقها على نطاق كبير في مجالات عديدة منها: مجال المال والأعمال والمجال الصحي ومجال التعليم وغيرها من المجالات,وهذا الأسلوب هو أسلوب التنقيب في البيانات. وقد هدفت الدراسة إلى التوصل إلى إمكانية الاستفادة من هذا الأسلوب في تطوير عملية المراجعة وبما يساعد على الاستفادة منه في شتى المجالات المحاسبية. وتوصلت الدراسة إلى أهمية هذا الأسلوب وإمكانية الاستفادة منه في المراجعة، حيث يحتوي على أدوات متعددة تعتبر كل منها أداة مفيدة للمراجع ويؤدى كل منها دوراً إيجابيًا في تطوير عملية المراجعة، وزيادة الكفاءة والفعالية.

٤/١/٢ دراسة (رفاعي, رهف الحريري , ٢٠١٧)

هدفت هذه الدراسة إلى الوصول إلى نموذج لدعم قرار تسعير التأمين التكميلي على المركبات من خلال الاعتماد على البيانات التاريخية والتنقيب فيها ، وذلك باستخدام تقنيتي شجرة القرار وشبكة بيرسيبترون العصبية متعددة الطبقات .حيث تم بناء نماذج شجرة القرار باستخدام خوارزمية C4.5 عن طريق برنامج WEKA، وذلك بالاعتماد على مجموعة من بيانات التدريب وقد تم اختبار نماذج شجرة القرار الناتجة من مرحلة التدريب من خلال مجموعة بيانات الاختبار باستخدام طريقة التحقق من الصحة المتقاطع ، وقد تم تفسير القيم الناتجة في حين تم تعليم نماذج شبكة بيرسيبترون العصبية متعددة الطبقات باستخدام خوارزمية الانتشار الخلفي للخطأ عن طريق برنامج WEKA وتم إختبار نماذج الشبكة الناتجة, و في النهاية تمت مقارنة نتائج اختبار جميع النماذج الناتجة لتحديد النموذج الأفضل , وتوصلت الدراسة إلى أن نموذج دعم قرار تسعير التأمين التكميلي على المركبات وفق عدد الحوادث وقيمة التعويض معاً والذي تم بناؤه باستخدام شجرة القرار كان له قدرة أكبر على التنبؤ وتفسير النتائج مقارنة بباقي النماذج التي تم بناؤها.

٥/١/٢ دراسة (سليمان, أسامه ربيع أمين, ٢٠١٩)

هدفت الدراسة إلى تحليل أهم محددات الاحتفاظ بالأقساط في تأمينات الممتلكات والمسئولية، وذلك من خلال تحليل العوامل والمتغيرات المؤثرة على مستوى الاحتفاظ بالأقساط في سوق التامين المصري في ضوء المحددات التأمينية والمحددات الاقتصادية؛بغرض بناء نموذج رياضي بالاعتماد على شجرة قرارات تصنيفية مبنية على خوارزمية C4.5 للتنبؤ بمستوى الاحتفاظ في سوق تأمينات الممتلكات والمسئولية 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 هي عدد شركات التأمين و الرافعة المالية لسوق التأمين، حيث إن 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 أيضاً على زيادة مستوى الاحتفاظ بالأقساط في سوق التأمين المصري. وأوصت الدراسة بضرورة زيادة الاعتماد من جانب

متخذي القرار في شركات التأمين على أساليب التنقيب في البيانات بصفة عامة، وشجرة القرارات التصنيفية على وجه الخصوص في التنبؤ بمستوى الاحتفاظ بالاقساط.

٦/١/٢ دراسة (علي، هدى عبدالرحيم حسين, ٢٠١٨)

تناول البحث التعريف بتقنية تنقيب البيانات وأهميتها، وقام الباحث باختبار عينة من المصارف الأهلية العراقية ممثلة بثلاثة مصارف، وتم جمع البيانات بالاعتماد على التقارير السنوية والحسابات الختامية الخاصة بالمصارف المبحوثة. وقد هدف البحث إلى اختبار فرضية رئيسية تنص على أن اعتماد خوارزمية CART يعد أفضل خوارزميات تنقيب البيانات لتحليل المؤشرات المالية للمصارف المبحوثة، وذلك للإجابة على تساؤلات مشكلة البحث والوصول إلى أهدافه الموضوعة. وتوصل البحث من خلال الفروض والأهداف إلى مجموعة من النتائج من بينها: تباين المؤشرات المالية للمصارف المبحوثة إذ جاء مصرف الشرق الأوسط متقدماً على بقية المصارف في تحليل مؤشرات أدائه. وأوصى الباحث بمجموعة من التوصيات أهمها: ضرورة إعطاء إدارات المصارف المبحوثة الإهتمام الكافي الذي يستحقه مفهوم تقنية تنقيب البيانات، والخوارزميات العاملة عليها للوصول إلى الأهداف المرجوة.

٧/١/٢ دراسة (مؤمن , شريف محمد لطفي, ٢٠١٩)

هدفت هذه الدراسة إلى التوصل لصياغة مدخل أو تصور مقترح لتطوير الممارسات الرقابية من خلال تقنية التعدين المالى للبيانات وذلك بعد إضافة إستراتيجية التوصية المعلنة لها، وتم تحديد المفهوم والأهداف والعقبات المتوقعة وخصائص نظم المعلومات المحاسبية فى ظله, وبينت الدراسة أنه في ظل التوجه العالمي نحو اقتصاديات المعرفة التي تعتمد بشكل أساسي على التقنيات الحديثة في استخدام المعرفة لتحقيق الرفاهية الإجتماعية واستثمار الموارد الإقتصادية المختلفة بشكل علمي ، أصبحت الإدارة الرقمية وسيلة بقاء وأداة لا يمكن الإستغناء عنها في عالم مفتوح على عنصر التغير والإبتكار والتنافسية التى تعد بمثابة معايير تعكس مستوى الأداء والنمو الاقتصادي لمنظمات الأعمال المعاصرة والمؤسسات الكبيرة الحجم والمتوسطة والصغيرة ، ليس فقط في زيادة حجم الإنتاج من السلع والخدمات، بل لأنها تعمل على زيادة كفاءة وفعالية الأداء التنظيمي لها وتحسين صورة مخرجاتها وتسريع عمليات تبادل معلوماتها عبر الشبكات، وهذا الأمر مكن منظمات الأعمال

المعاصرة من الاستفادة بشكل كبير من مزايا هذه الإدارة لغرض إعادة تصميم وتشكيل بنيتها التحتية, و يفرض التطور الرقمى الحادث فى مجال إدارة الأعمال التجارية فى العالم الآن ضرورة التوسع فى الكشف عن تقنيات حديثة لدعم الممارسات الرقابية فى منظمات الأعمال عند التحول من المعالجات اليدوية للبيانات إلى معالجات رقمية بالكامل، خاصة بعد أن فرضت تكنولوجيا المعلومات نفسها على النظم المحاسبية وحولت جانب كبير منها للنظام الرقمي لتكون فى صورة تقارير وقوائم مالية يتم بثها عبر قواعد البيانات المجمعة أو مستودعات البيانات، أو من خلال الشبكة الدولية للمعلومات (الإنترنت)، حيث أصبحت المعاملات المالية للمنظمات الآن تعمل من خلال ثلاثة إتجاهات ، الإتجاه الأول إتجاه تكنولوجى رقمي والثاني إتجاه بشرى إجتماعى، والأخير إتجاه يدوي تقليدى للعمليات التجارية ما ازل يعمل وفقا للنماذج الورقية التقليدية.

٨/١/٢ دراسة (نائب, إبراهيم عبدالواحد, ٢٠١٩)

هدفت الدراسة إلى حل مشكلة إختيار الجهة المنفذة للمشاريع الإنشائية (مشاريع طرق– مشاريع أبنية مدرسية) باستخدام التنقيب في بيانات المشاريع الإنشائية لمديرية الخدمات الفنية بحلب, وذلك باستخدام أشجار القرار لتقييم المشاريع ولدعم قرارات مديرية الخدمات الفنية بحلب لاختيار الجهة المنفذة للمشاريع المستقبلية من خلال استخدام منهجية التصنيف معتمدة على خوارزمية CART في عملية بناء النموذج الشجري باستخدام أداةWeka واتباع طريقة Cross– validation– 10 في عملية التدريب والاختبار، وتم تطبيق هذه الدراسة على بيانات المشاريع (التدريب والاختبار) كمرحلة أولى، واختيار الجهة المنفذة للمشاريع مرحلة ثانية (دالة الهدف) ، وتم بناء نموذج تم تقييمه وكانت نتائجه جيدة وهو قابل للتطبيق العملي وقابل للتعميم, وقد أوصت الدراسة بضرورة استخدام هذا النموذج في إختيار الجهة المنفذة للمشاريع المستقبلية في مديرية الخدمات الفنية بحلب, وتناولت الدراسة تحديد المعايير اللازمة لاتخاذ القرارات في الجمعيات الخيرية في سوريا, وبالتالي إيجاد آلية بسيطة لتسهيل عملية الخاذ القرارات,وقد اعتمدت الدراسة على المنهج الوصفي وهدفت إلى تحديد المعايير ودراستها وتحليلها لتصنيف العائلات بحسب الأولية، ودرجة الاحتياج وذلك باستخدام خوارزمية C4.5 ، وقد تم استخدام برنامج WEKA في بناء النموذج لتقييم العائلات،وتم بناء شجرة القرار بالاعتماد على مجموعة من بيانات حجمها كبير، وتقييم نموذج شجرة القرار من خلال تفسير دقيق للقيم الناتجة. وتوصل البحث إلى إمكانيّة بناء نموذج شجرة قرار تصنيف العائلات ذات دقة عالية إذ بلغت دقة النموذج باستخدام خوارزمية C4.5 حوالي(0.9691) وهي نسبة عالية جداً.

وأوصت الدراسةبأهمية تطوير عملية اتخاذ القرار باستخدام تقنيات إضافية مثل المنطق الضبابي لإزالة ضبابية القرار والوصول إلى قرار واضح خالي من الغموض.

٩/١/٢ دراسة (إبراهيم, أحمد اسماعيل محمد وآخرون, ٢٠٢٠)

هدفت هذه الدراسة إلى تحليل وتقييم حجم المستفيدين من برنامج الصندوق القومي للتأمين الصحي علي صعيد المستوي الاقليمي(الولايات) باستخدام اساليب التنقيب عن البيانات الحديثه في السودان, وذلك لخلق مؤشرات تساعد علي تنميه الصندوق وتغطيه اكبر لشرائح المجتمع المختلفه , وقد تم الحصول البيانات اللازمة من نوع SQL لسنه ٢٠١٩ وتم إجراء عمليات التحضير اللازمه من تنظيف وتجهيز واعادة صياغه عليها بغرض تجهيزها لعمليات التنقيب اللازمة وذلك للحصول علي أنواع وأنماط جديدة تساعد في إتخاذ القرارت والخطط المستقبليه للصندوق القومي للتامين الصحي, وفي هذا الصدد تم استخدام خوارزميه decision tree و J48 للحصول علي قراءات وتصنيفات من البيانات الخاصه بالمستفيدين من التامين الصحي وتوصلت الدراسة إلى العديد من النتائج أهمها أن ولايه الجزيرة هي أكثر الولايات المستفيدة من برنامج الصندوق وأن أكثر القطاعات المستفيده هو قطاع الأسر الفقيرة , وكذلك تبين وجود قصور في قطاع المحاميين وأتضح أن عدد الإناث أكثر من عدد الذكور نسبياً , وأن هناك قصور في ولايات غرب ووسط دارفور, تم استخدام برنامج SQL SERVER 2013 في قواعد البيانات وتم تنفيذ خوارزميات التصنيف والتنبؤ عن طريق برنامج . weeka .

٢/٢ الدراسات الأجنبية

١/٢/٢ دراسة (Bhowmik,2011)

تناولت الدراسة طرق الكشف عن الاحتيال في تأمين السيارات؛ وذلك باستخدام تقنيات التنقيب في البيانات, وقدمت الدراسة طريقة للكشف عن الاحتيال في تأمين السيارات، وتحليل أنماط الاحتيال في البيانات؛ وذلك باستخدام مصنف بييز المبسط Naïve Bayesian Classification، وخوارزميات معتمدة على أشجار القرارات مثل خوارزمية C4.5, وتوصلت الدراسة إلى تقديم وصفاً موجزاً للخوارزميات مع تطبيقاتها للكشف عن التزوير والغش في التأمين على السيارات, وهذا بالاضافة إلى تقييم النماذج الناتجة.

٢/٢/٢ دراسة (Jozebencna, 2011)

قامت هذه الدراسة باستخدام منهج التصنيف والتحليل في استخدام أشجار القرار والمنطق الضبابي, وقد تم تطبيقها لدراسة حالة الاستثمار وتقييم المشاريع ضمن قسم البلدية في سلوفينيا، بالإضافة لتقييم كفاءة وفعاليّة المؤسسات العامة؛ وذلك بإيجاد حلول لها خصائص معدّلة بشكل معين للإستخدام في القطاع العام، وتهدف هذه الدراسة إلى جعل تطبيقات التقييم تقدم تقييماً دقيقاً وبشكل أفضل،وتكون سهلة الوصول والاستخدام، ويمكن تطبيق النموذج للتقييم في كافة المجالات ضمن القطاعين العام والخاص، وتم استخدام مكتبة برمجة خاصة لبناء النموذج وكانت نتائج البحث دقيقة وجيدة لمساعدة متخذي القرار للحصول على قرار دقيق وواضح.

٣/٢/٢ دراسة (Akinsola Adeniyi et al.,2015)

قامت الدراسة ببناء شجرة قرار للتنبؤ بمستوى الخطر في تأمين المركبات؛ وذلك بالاعتماد على خوارزمية ID3 وباستخدام بعض المتغيرات المستقلة, وهدفت الدراسة إلى الوصول لنموذج له القدرة على التنبؤ بمستوى الخطر, وتوصلت الدراسة إلى إمكانية استخدام خوارزمية ID3 للوصول إلى هذا النموذج, وبعد تطبيقه تبين الحصول على نتائج مرضية, وقد أوصت الدراسة بضرورة استخدام تقنية شجرة القرار للتنبؤ بالأخطار ومستوياتها في تأمين المركبات.

٤/٢/٢ دراسة (Vaidehi R,2016)

تناولت الدراسة البيانات الفعلية المقدمة من آلة UCI عن طريق البيانات المقدمة من أحد البنوك البرتغالية. وهدفت الدراسة إلى التوصل إلى نهج مقترح لتعزيز العملية التنبؤية من خلال المقارنة بين ثلاث تقنيات هي: النموذج الخطي العام، وأشجار القرار من خلال خوارزمية C4.5، والنمذجة القاعدية. وقد توصلت الدراسة إلى تحديد ٦ متغيرات هي الأفضل في بناء النموذج,كما توصلت إلى أن نموذج المصنفات القاعدية هي الأفضل في التطبيق.

٥/٢/٢ دراسة (DOSKOCIL, 2016)

هدفت الدراسة إلى الوصول إلى نموذج ضبابي خبير لاتخاذ القرارات؛ وذلك لتقييم المخاطر الكلية لأي مشروع, وقد تم بناء هذا النموذج الضبابي على طريقة (RIPRAN (Risks Project Analysis وهي عبارة عن طريقة تجريبية لتحليل مخاطر المشروع، حيث تم استخدام المنطق الضبابي في برنامج MATLABلإنشاء نموذج ضبابي لاتخاذ القرار، وتبين أن الميزة في هذا النموذج الضبابي تكمن في القدرة على تحويل متغيرات الدخل إلى متغيرات لغوية، بالإضافة إلى التقييم اللغوي للقيمة الإجمالية لمخاطر المشروع ومن الممكن محاكاة قيمة المخاطر وعدم التأكد المرتبطين بها.

٦/١/٢ دراسة (Werner,M. , 2017)

ركزت هذه الدراسة على تقنية التنقيب المالى للبيانات، وتتمثل فى ترتيب الأحداث المالية لزيادة فاعلية المراجعة الخارجية على التدفقات النقدية، وقد ركزت هذه التقنية على ترتيب الأحداث والمعاملات المالية بدلاً من الترتيب الزمنى لهذه المعاملات والأحداث المالية، وذلك لزيادة كفاءة ودقة عملية المراجعة الخارجية على تلك الصفقات أو الأحداث المالية,وقد بينت الدراسة وجود منهجا جديداً في مجال التنقيب المالي للبيانات بصفة خاصة متمثلة في ثلاثة إستراتيجيات أو أسس بهدف أغراض التقييم, وتعد هى الدراسة التى جمعت بين الإستراتيجيات الثلاثة لتقنيات التنقيب المالي للبيانات من الدراسات التي توصلت لوجود علاقة بين تلك الإستراتيجيات الثلاثة , وعلى الرغم من ذلك فإنها لم تقم بتحديد أية من الأدوات المستحدثة فى مجال التنقيب المالي للبيانات والتى تعمل على أساس هذه الإستراتيجيات .

٧/٢/٢ دراسة (Esraa H. Abd Al-Ameer, 2020)

هدفت الدراسة إلى تصنيف الوثائق باعتباره أحد أهم المجالات في معالجة اللغات الطبيعية وتنقيب النصوص, وأوضحت أن هناك العديد من الخوارزميات التي يمكن استخدامها لعمل هذا التصنيف، وركزت على تحسين TC عن طريق اختيار الخواص, وهذا يعني تحديد مجموعة فرعية من الخواص الأصلية دون التأثير على دقة العمل، حيث اقترحت طريقة جديدة لاختيار الخواص والتي يمكن أن تكون صيغة عامة ونموذج رياضي (RFE) وذلك عن طريق مقارنة الطريقة المستخدمة مع طريقتين من الطرق الشائعة لاختيار الخواص وهما طريقة Chi-square وطريقة Threshold, وأثبتت نتائج الدراسة أن الطريقة الجديدة منافسة للطرق الأخرى المعروفة، وكانت

أفضل النتائج٨٣٪ عند استخدام٦٠ %من الخواص، و ٨٢ %عند استخدام ٤٠ %من الخواص، و ٨٢ %عند استخدام ٢٠ %من الخواص, وقد تم إجراء الاختبارات باستخدام خوارزميتي التصنيف Decision Tree (DT)و Naïve Bayes (NB)حيث مجموعة البيانات المستخدمة هي مجموعة بيانات معروفة (٢٠ newsgroups text) تتكون من (١٨٨٤٦) مستند,وتوصلت الدراسة إلى عدة نتائج أهمها:أن طريقة اختيار الخواص المقترحة قابلة للمقارنة مع الطريقة القياسية لاختيار الخواص مثل(Chi-square) .

٣- تقنية التنقيب في البيانات

إذا افترضنا أن إحدى شركات التأمين لديها كمية هائلة من البيانات المتعلقة بفرع معين من فروع التأمين, وأن ما نبحث عنه بداخلها هي البيانات الثمينة بينما الباقي هو مجرد نفايات لا فائدة منها,ففي هذه الحالة تظهر أهمية التنقيب في البيانات (Data Mining).

١/٣ مفهوم التنقيب في البيانات:

يمكن تعريف تقنية تنقيب البيانات على أنها إحدى التقنيات الهامة والمسئولة عن تحصيل وانتاج المعلومات وإدارة المعرفة من خلال التعامل مع كم كبير من البيانات (البيانات الكبيرة)،والتي يتم التنقيب فيها وتحليلها بأساليب علمية حديثة؛ وذلك من أجل الوصول إلى معرفة ومعلومة جديدة، ومن هنا يمكن تعريفها على أنها عملية تحليلية لتحويل البيانات إلى معلومات يمكن استخدامهالزيادة أداء المنظمة من خلال بناء نماذج رياضية تساعدعلى اتخاذ قرارات أفضل(Saarenvirla,2001),ومن هذا المنطلق يمكن القول أن تقنية التنقيب في البيانات يمثل عملية متطورة تقومباشتقاق البيانات المطلوبة من بين كم هائل من البيانات طبقاً لأهدافتم تحديدها مسبقاً,ولذا فهي تعد بمثابة أداةتنبؤ مسبق،إذ تنقب البيانات ذاتياعن العلاقاتالمحتملة والشاذة لتحدد المشكلات التي لم يتم تحديدها من قِبل المستخدم(علي, هدى عبدالرحيم حسين ٢٠١٨).

إن تقنية التنقيب في البيانات عبارة عن منهجية تجمع بين نتائج الأبحاث في الذكاءالاصطناعي والفهم الآلي والتعرف على الأشكال وقواعد المعلومات والإحصاء،ولذلك فهي عبارة عن الاستكشاف الآلي لأنماط شائعة ومخفية في قواعد بيانات معينة, ومما سبق يمكن القول أن تقنية تنقيب البيانات تمثل ذلك الجزء من عملية اكتشاف المعرفة وإيجاد العلاقات بين البيانات، والتي تتم بواسطة طرق بإمكانها تكوين نماذج من البيانات ومن خلال تفسير وحساب المعلومات يمكن ظهور معرفة جديدة

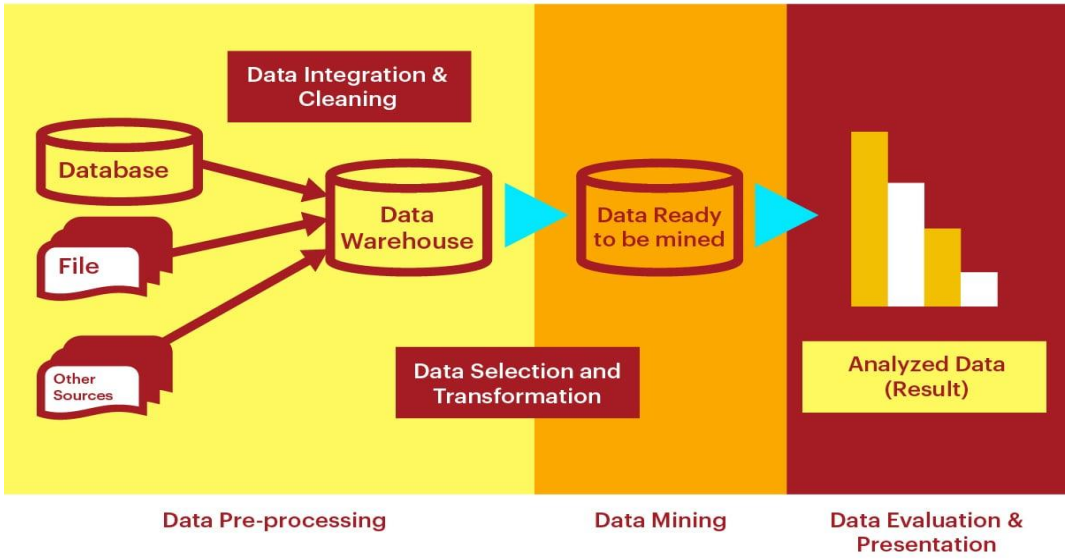
لصانعي القرار للعمل على أساسها في عدة مجالات مثل المحاسبة،الاتصالات،الخدمات الطبية، الخدمات المالية والمصرفية، والتأمين وغيرها، وبعد التطور الكبير في استخدام البيانات أصبح من الضروري إيجاد خوارزميات جديدة لخدمة البيانات، وايجاد أفضل تلك الخوارزميات للحصول على أحسن النتائج بين البيانات كمحاولة لوضع خصائص لتلك البيانات(رقمية،رمزية,مختلطة) ث ممطابقة تلك الخوارزميات مع خصائص البيانات، وهذه المعلومات تساعد في" تنقيب البيانات " لصنع قرارات ذكية في اختيار الخوارزمية الملائمة لملفات البيانات.

ومما تقدم يمكن القول أن تنقيب البيانات يشمل أربعة أوجه أساسية تتمثل فيما يلي (Karloff,2009):

- البيانات: وهي المادة الخام والمتمثلة بالحقائق والنصوص التي تتطلب المعالجة.
- المعلومات: وهي النماذج والعلاقات بين تلك البيانات التي تشكل معلومات مفيدة.
- المعرفة: وهي عبارة عن المعلومات السابقة، والتي تم تحويلها إلى معرفة حول الأنماط التاريخية أوالتوقعات المستقبلية.
- مستودعات البيانات: وهي عبارة عن قواعد البيانات المستخدمة في التحليلات الزمنية واكتشاف المعرفة واتخاذ القرارات.

والشكل التالي يوضح فكرة تخزين البيانات في مستودع بيانات.

شكل (١/٣) : كيفية تخزين البيانات في مستودع بياناتData Warehouse

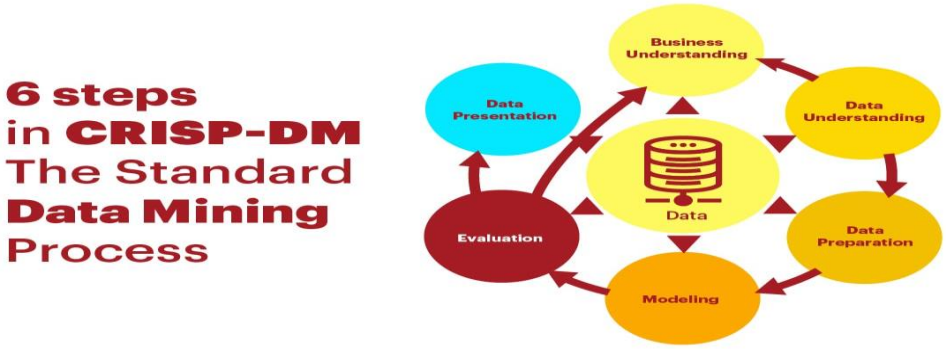


المصدر: (Matthew Anyanwu N. and Sajjan G. Shiva,2010)

٢/٣مراحل عملية التنقيب في البيانات القياسية:

هناك ستة مراحل أو خطوات تمر بها عملية التنقيب في البيانات القياسية(Crisp-DM) يمكن إيجازها فيما يلي:

شكل (٢/٣) : خطوات عملية التنقيب في البيانات



المصدر: (Matthew Anyanwu N. and Sajjan G. Shiva,2010)

١/٢/٣ فهم الأعمال: حيث تتم عملية التنقيب بناءً على تحديد مايتم البحث عنه فعلاً؛ لذا فانه لا بد من فهم

وتحديد متطلبات العمل من أجل صياغة المشكلة وذلك بمجرد تحديدها, ثم تحويل هذا الفهم إلى مشكلة تنقيب عن البيانات, وإعداد استراتيجية أولية لتحقيق تلك الأهداف.

٢/٢/٣فهم البيانات: وفي هذه الخطوة يتم تجميع البيانات الأولية وفهمها ثم متابعة أنشطتها؛وذلك لتحديد جودتها وتحديد مصدر تلك البيانات وكيف تم تجميعها, وهنا يتم استخدام تحليل البيانات الاستكشافي للتعرف على البيانات وأخذ فكرة أولية عنها وتقييمها.

٣/٢/٣إعداد أو تحضير البيانات: وهنا يتم تغطية جميع الأنشطة التي تم القيام بها لإنشاء مجموعة البياناتالنهائية من البيانات الأولية, ويجب ملاحظة أن تحويل البيانات الأولية إلى مجموعة بيانات تحليلية يشكل ٩٠٪ من وقت المشروع.

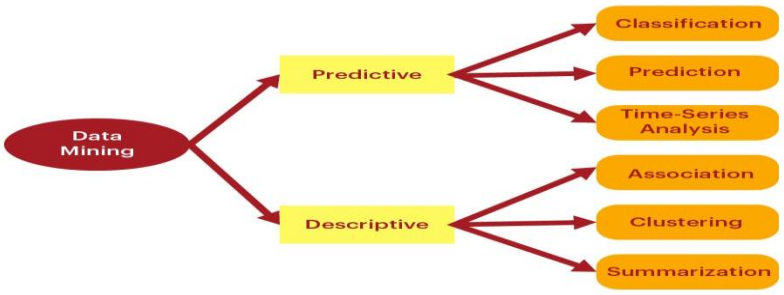
٤/٢/٣ النمذجة:والنماذج هي تطبيقات خوارزمية للبحث وتحديد وعرض أي أنماط أو رسالة في البيانات.

وهناك نوعان من النماذج في التنقيب عن البيانات هما:

- نماذج تصنيفية أو وصفية (Descriptive).
- نماذج تنبؤية (Predictive).

ويمكن تلخيص هذه النماذج في الشكل الآتي:

شكل (٣/٣) نماذج التنقيب في البيانات(DM)



المصدر: (Matthew Anyanwu N. and Sajjan G. Shiva,2010)

ويتم في هذه الخطوة (خطوة النمذجة) اختيار وتطبيق تقنيات النمذجة المختلفة، كما يتم تحديد المعلمات

(Parameters) الخاصة بها تجاه القيم المُثلى, ويجب ملاحظة أن هناك العديد من التقنيات التى تستخدم للمشكلة الواحدة في التنقيب عن البيانات

٥/٢/٣ التقييم: وفي هذه الخطوة يتم تقييم النتائج في سياق أهداف العمل,بهدف تحديد ما إذا كان هناك أي بيانات مهمة ليتم النظر فيها بشكل كافٍ,وفي نهاية هذه الخطوة يجب اتخاذ قرار الانتقال أو عدم الانتقال إلى مرحلة النشر.

٦/٢/٣ النشر :وفي هذه الخطوة نحتاج إلى تحديد كيفية استخدام النتائج,ويجب تنظيم المعرفة المكتسبة وعرضها بطريقة يمكن لمتخذي القرار استخدامها وفقًا لمتطلباتهم، ومرحلة النشر يمكن أن تكون بسيطة مثل إنشاء تقرير، أو معقدة مثل تنفيذ عملية تنقيب في البيانات القابلة للتكرار عبر قاعدة البيانات.

٣/٣ أهم أنواع تقنيات التنقيب في البيانات(Atre,Shaku,2001)

هناك العديد من التقنيات المستخدمة في التنقيب في البيانات أهمها:

١/٣/٣ تقنية التحليل العنقودى Clustering Analysis

وتقوم هذه التقنية على جمع كميات كبيرة من البيانات معًا اعتمادًا على أوجه التشابه في المجموعات أو ما يسمى العنقودية, وتعمل هذه التقنية على تحليل المجموعات بعدة طرق بما في ذلك تجميع البيانات المتشابهة التي تساعد في فهم الهيكل الداخلي للبيانات، واكتشاف المعرفة للبيانات،وتظهر أهمية هذه التقنية في استكشاف البيانات، واكتشاف العيوب، ولعل خوارزمية مجموعة الوسائل k والوسائل C وتعظيم التوقعات(EM) من أهم الخوارزميات الشائعة لهذه التقنية.

٢/٣/٣ تقنية تحليل قواعد الارتباط Association Rule Analysis

إن عملية التنقيب الخاصة والمرتبطة تهدف إلى مراقبة الأنماط المتكررة أو الارتباطات من مجموعات البيانات, وهي إحدى تقنيات التنقيب عن البيانات غير الخاضعة للإشراف، والتي يتم فيها تعريف مجموعة العناصر على أنها مجموعة من عنصر واحد أو أكثر.

وهى في الأساس تقنية تعلم آلي قائمة على القواعد يتم استخدامها لاكتشاف العلاقات بين المتغيرات في مجموعات البيانات باستخدام (If / Then) بما يتضمن جزئين رئيسيين وهما: مقدمات ونتائج,وتتمثل إحدى مميزات هذه التقنية في أنها تمر بعدد قليل من قواعد البيانات أثناء البحث في

مساحة فردية،ولعل ذلك مهم جداً لحل بعض المشاكل مثل تحليل سلوك العملاء,وتعد خوارزمية APRIORIو SETM و Eclat وغيرها من أشهر خوارزميات تقنية تحليل قواعد الارتباط.

٣/٣/٣ تقنية تحليل الإنحدار Regression

وتعني أنه يمكن رسم خط مستقيم لإظهار كيفية ارتباط كل متغير مع المتغيرات الأخرى خاصة إذا كان التحليل يهدف إلى إجراء تنبؤ استنادًا إلى تأثير أحد المتغيرات على المتغيرات الأخرى.

٤/٣/٣ تقنية التصنيف Classification

والتصنيف هو أسلوب شائع للتنقيب عن البيانات يُشار إليه على أنه أسلوب تعلم خاضع للإشراف حيث يتم من خلاله استخدام مجموعة بيانات نموذجية لتعلم بناء المجموعات, ومن خلال هذه التقنية يتم بناء مجموعاتلبيانات من الأمثلة مقسمة بالفعل إلى مجموعات يشار إليها بالفئات,كما يتم تحقيق تعلم هذه الفئات عادةً باستخدام نموذج يُستخدم لتقدير معلمات المجموعة والمعروف أيضًا باسم تسميات الفصل الواحد,وتشمل بعض تطبيقاته التسويق المستهدف للعملاء، وتصنيف المستندات، وإدارة الأمراض الطبية، وتحليل بيانات الوسائط المتعددة وغيرها من المجالات المختلفة.

٥/٣/٣ تقنية تحليل شجرة القرارات Decision Tree

شجرة القرارات Decision Tree تُعد من أهم النماذج التنبؤية للتنقيب عن البيانات، وهى أيضاً من أدوات دعم إتخاذ القرار. وتعتمد شجرة القرارات على تكوين خريطة تبدأ بالملاحظات وتنتهى بالنتائج المحتملة،وتتكون شجرة القرارات من سلسلة من العقد والأفرع، وتمثل كل عقدة فى الشجرة إختبار عن إتجاه، بينما يمثل كل فرع من فروع شجرة القرار النتائج المحتملة، وبإستخدام نموذج شجرة القرارات فى صنع القرارات يمكن البحث من الجذر إلى الأوراق فتحتوى عقد الأوراق على نتائج كل تصنيف (R,Bhartiya.,p,Sharma,2012),كما تُستخدم شجرة القرارت فى الكشف عن غش بطاقات الإئتمان وتأمين السيارات وغش الشركات، ومن ثم يمكن إستخدامها فى التنبؤ بالمطالبات والتعويضات في فروع التأمين المختلفة,وتعد هذه التقنية من أشهر التقنيات المستخدمة في عمليات التنقيب عن البيانات وأكثرها انتشاراً، وهي طريقة شائعة لاتخاذ القرارات الهامة,وتعد تقنية تحليل شجرة القرارات خوارزمية تعلم آلي خاضعة للإشراف تركز على نمذجة المدخلات، وكذلك علاقات الإخراج في شكل قواعد If / Then وهي تساعد في الكشف عن البيانات الاحتيالية، وإدارة الأعمال، وإدارة علاقات العملاء، وتشخيص الأخطاء, وجدير بالذكر أن هناك نوعان من تحليلات شجرة القرارات المستخدمة في عملية التنبؤ النوع الأول هو التصنيف(Classification) ويعتمد على المنطق

واستخدام مجموعة متنوعة من الشروط مثل (إذا/ أو) ,(نعم / لا) ، حتى يتم تحديد جميع البيانات ذات الصلة, والنوع الآخر هو الإنحدار(Regression) وهذا النوع يستخدم عندما يكون القرار المستهدف قيمة رقمية, وهناك أنواع من تحليلات شجرة القرارات تستخدم في عملية التوصيف لأنماط البيانات مثل تحليل الرابط (Link Analysis)، والبحث عن الانحراف (Deviation Detection)(Quinlan,J. R. 1993).

٤/٣ أهمية تقنية التنقيب في البيانات

لقد ظهرت أهمية تقنية تنقيب البيانات بوضوح نتيجة التطور الكبير الذي شهده استخدام قواعد البيانات في النصف الثاني من تسعينات القرن العشرين، وفي نفس الوقت ظهرت أيضاً بوضوح الحاجة إلى ما يسمى "اكتشاف المعرفة"، والذي أتاح التعامل مع كمية كبيرة من البيانات(Big data), ومن هنا يمكن توضيح أهمية تقنية تنقيب البيانات وتوفير معلومات دقيقة وصحيحة من خلال النقاط التالية(Thuraisingham,et al.,2008,).

١/٤/٣ توفير معلومات دقيقة وصحيحة من خلال إتاحة التعامل مع تقنيات المعلومات المتطورة، وبما يؤمن قياس فاعلية وانتاجية نُظم المعلومات الفرعية المختلفة.

٢/٤/٣ تُساعد في الاستخدام الفعَال لمصادر البيانات والموارد المتاحة.

٣/٤/٣ تشجع المنظمات على النمو ومواكبة التطورات والمستجدات في مختلف مواردها المادية والفنية والبشرية ،ومن ثم زيادة المعرفة والحد من البدائل وتقليص حالة عدم التأكد المرتبطة بها.

٤/٤/٣ تُمكن مُتخذي القرار من إجراء عملية التخطيط واستمراريته بما يساهم في تحديد المشكلة وعناصرها واتمام الرقابة على ملفات المعلومات والبيانات اللازمة لاتخاذ قراراتهم.

٥/٤/٣ تساعد مُتخذي القرار فيما يحتاجون إليه من معلومات لاتخاذ قراراتهم كما تُساهم في سرعة إنجاز العمل والمهام ،وتبسيط الإجراءات (Nisbet et al., 2009).

٦/٤/٣ استخلاص المعلومات المخبأة في البيانات بما يوفر للمنظمات القدرة على الاستكشاف في مختلف الأنشطة وبناء التنبؤات والتكهن بالسلوكيات والاتجاهات المستقبلية؛ بما يسمح بتقدير القرارات الصحيحة واتخاذها في الوقت المناسب.

٧/٤/٣ تستخدم في المجالات المختلفة كما في مجال المال والأعمال للتنبؤ بالربح والخسارة، ومن ثم الإجابة على العديد من التساؤلات وفي وقت قياسي، وخاصة تلك الأسئلة التي من الصعب الإجابة

عليها إن لم يكن مستحيلاً باستخدام تقنيات الإحصاء التقليدية، والتي إن وجدت فإنها تستغرق وقتاً طويلاً وإجراءات تحليل معقدة.

٨/٤/٣ تعمل على مجاراة التحولات والتطوارت في كافة الميادين بما يمكن من الاستكشاف والاستنباط والتنبؤ والوصول إلى قرارات أفضل وبالسرعة المطلوبة.

٤- تأمين البترول في شركات التأمين في السوق المصري

تتوفر في شركات التأمين المصرية خدمات تأمين أنشطة البترول والغاز الطبيعي والطاقة للتغلب على التحديات السلبية المتعلقة بمخاطر استكشاف، وإنتاج النفط والغاز،وتدعيم وتطور التقنيات، وتشجيع الاستثمار في مجالات الطاقة البديلة مثل: طاقة الرياح والطاقة الشمسية، ومن أهم البرامج التأمينية التي توفرها شركات التأمين ما يلي:

- برامج خاصة لجميع الممتلكات البرية والبحرية مثل: تسهيلات إنتاج البترول والغازومحطات التكريرومصانع البتروكيماويات أثناء مراحل الإنشاء والتشغيل ضد جميع الأخطار المادية، بالإضافة إلي توقف الأعمال وفقد الإيراد.
- تأمين الآبار البرية والبحرية ضد أخطار الحريق والانفجار والخروج عن السيطرة.
- تأمين المسئوليات المدنية والأخطار الناتجة عن التلوث.
- تأمين الإرهاب، والتخريب، والإضرابات الأهلية والعمالية، والمخاطر السياسية.
- برامج خاصة بتأمين الحفارات البرية.
- برامج خاصة بتأمين الحفارات البحرية، بالإضافة لبرامج أندية الحماية والتعويض.

وعلى الرغم من وجود ٢٥ شركة تأمين تعمل في نشاط تأمين الممتلكات والمسئولية على مستوى السوق المصرية في مارس ٢٠٢١،إلا أن هناك ٦ شركات منها فقط هي التي توفر تغطيات تأمينية للمخاطر التي تواجه قطاع البترول وتضم قائمة الشركات الستة كل من:شركة مصر للتأمين، وشركة قناة السويس للتأمين، وشركة جي آي جي"GIG" ، والمصرية للتأمين التكافلي ممتلكات، وأورينت للتأمين التكافلي ممتلكات، وشركة تشب للتأمين (الكتاب الإحصائي السنوي عن نشاط سوق التأمين في مصر, الهيئة العامة للرقابة المالية ,سنوات مختلفة).

ولقد نجحت الشركات الستة في زيادة حصيلة أقساطها المباشرة لتصل إلي ١,٩ مليار جنيه عام ٢٠٢٠ مقابل ١,٥ مليار جنيه تقريباً في عام ٢٠١٨ ، وبينما لم تتجاوز الأقساط ١,١ مليار جنيه في ٢٠١٧(الكتاب الإحصائي السنوي عن نشاط سوق التأمين في مصر, الهيئة العامة للرقابة المالية,٢٠٢٠)

١/٤ أنواع وثائق تأمينات البترول والحفر والتغطيات الإضافية

تعد مراحل الإنتاج هى الأساس فى عمل التغطيات التأمينية لنشاط البترول، حيث تغطيات ترتبط بنشاط البحث أو التنقيب أو الاستكشاف، مثل تأمين أجهزة الحفر البرية والبحرية، وتأمين حفر الآبار فى البر والبحر"وثائق متخصصة"، ووثائق ترتبط بمشروعات إنشاء تسهيلات إنتاج البترول براً أو بحراً، مثل تأمين مشروعات إنشاء خطوط الأنابيب براً وبحراً وإنشاء أرصفة أو منصات الإنتاج البحرى وتركيب أجهزة الإنتاج لاستخراج الزيت الخام من الآبار، ثم إنشاءات تتعلق بأجهزة المعالجة الأولية حيث يتم إزالة والتخلص من الأملاح والكبريت والماء والغازات المصاحبة للبترول " تغطيات تقترب من التأمين الهندسى".

كما يوجد تغطيات ترتبط بمرحلة بدء الإنتاج والاستخراج من الآبار براً أو بحراً بعد إنتهاء عمليات الإنشاء وهى وثائق مرحلة التشغيل مثل: وثيقة جميع الأخطار المادية، ووثيقة تأمين تكاليف التحكم فى الآبار، وفيما يتعلق بالتغطيات التى ترتبط بشركات خدمات البترول، والتى تقدم خدمات كثيرة ومتنوعة لشركات الانتاج، توجد عدة أمثلة منها خدمات صيانة أجهزة الاستخراج، وصيانة الآبار، وخدمات قياس المخزون، وتوريد الطفلة والأجهزة المساعدة أثناء الحفر، وفحص وإختبار المواسير، وخدمات الأسمنت وتمهيد الطرق للحفار وتأجير الوحدات البحرية وغيرها من خدمات كثيرة.

ويوجد أيضاً وثائق ترتبط بمرحلة نقل الزيت الخام والغاز المسال LNG داخلياً أو بين الدول، ويستخدم فى ذلك وثائق تأمين شحنات بترول بحرى متخصصة لنقل الزيت الخام بحراً عن طريق Tankers أو ما يسمى LNG Ships وتنقل الغاز المسال بالتانكات الكروية إلى الدول المستوردة (وثائق مرتبطة بتأمين الشحنات البحرية) وقد يصاحب البترول خروج غازات وتسمى غازات بترولية مثل غاز البروبان والبيوتان (البوتاجاز) وهى تختلف فى تركيبها واستعمالاتها، وكان قديما –وقبل اكتشاف أهمية الغاز– يتم التخلص من الغازات البترولية عن طريق خط أنابيب يتم إشعاله للتخلص من الغاز نراها شعلات فى البر والبحر وتسمى Flare.

ويتم تأمين كافة آلات مواتير الرفع، والوايرات الحديدية، وجسم الحفار، وموتور التدوير، والقائم المعدنى الذى ينتهى ببلورات الحفر (ذراع الحفر)، ومانع الانفجار BOP والذى يشترط تثبيته على الحفار للتحكم ومقاومة كمية الضغط الناتج من الحفر والقادمة من باطن الأرض ومواسير الطفلة وتانك الطفلة.

وتستثنى الأجزاء التى تعمل بشكل كامل تحت سطح الارض وقد تلتصق بعض تلك الأجزاء بالصخور أثناء الحفر دون وقوع حادث انفجار ويصعب استخراجها، وقد تسقط بالبئر نتيجة انفصالها عن ساق الحفر وفى كلا الحالتين لا يتم تعويضها منفردة، ولكن فى حالة الانفجار أو الحريق يتم تعويض تلك الأجزاء بالكامل، كما يستثنى خطر الإرهاب والتخريب والعنف السياسى والحروب والاستيلاء على السلطة والحرب النووية وغيرها.

وتختلف أنواع وثائق تأمينات البترول وفقاً لمراحل الإنتاج، ويمكن توضيح ذلك على النحو التالي:

١/١/٤وثائق ترتبط بنشاط البحث أو التنقيب أو الإستكشاف مثل: تأمين أجهزة الحفر البرية والبحرية، وتأمين حفر الآبار في البر والبحر.

٢/١/٤ وثائق تتعلق بمشروعات إنشاء تسهيلات إنتاج البترول في البر والبحر مثل: تأمين مشروعات إنشاء خطوط الأنابيب، وإنشاء أرصفة أو منصات الإنتاج البحري، وتركيب أجهزة الإنتاج لاستخراج الزيت الخام من الآبار.

٣/١/٤وثائق تتعلق بالإنشاءات وبأجهزة المعالجة:وهي تلك الوثائق التي ترتبط بمرحلة بدء الإنتاج والاستخراج من الآبار بعد انتهاء عمليات الإنشاء، وهي وثائق مرحلة التشغيل مثل: وثيقة جميع الأخطار المادية، ووثيقة تأمين تكاليف التحكم في الآبار Cost Of Well Control Policy ، وتختلف هذه الوثائق عن بعضها البعض فإذا كان الإنتاج للتصدير دون استخراج مشتقاتمعالجة أولية فقط، فتقتصر الوثيقة علي تأمين أجهزة الإنتاج المثبتة علي البئر، وخطوط الأنابيب وتانكات التخزين وأجهزة إزالة الملوحة والماء ومعالجة الغازات، أما في حالة معالجة الزيت الخام لاستخراج مشتقاته فهناك وثائق تأمين منفصلة لتغطية ما يسمي بمصفاة تكرير البترول، والتى تنتج بنزين الطائرات Jet Fuelوالبنزين عالى الأوكتين ٩٥ و ٩٢ و ٨٠، ثم الديزل والنافتا والكيروسين، والمازوت الذى يستخدم فى أعمال رصف الطرق، أو تشغيل بعض محطات الكهرباءأولية؛ حيث يتم إزالة والتخلص من الأملاح والكبريت والماء والغازات المصاحبة للبترول.

٤/١/٤ وثائق تأمين صناعة البتروكيماويات: حيث قد يصاحب البترول خروج غازات بترولية ، وبالنسبة لوحدات الحفر البرية، فيتم التأمين عليها وفق شروط All Risks Oil & Gas Drilling Equipment Form (land).

٥/١/٤ التأمين على وحدات الحفر: ويتم التأمين على وحدات الحفر وفقاً لنموذج تأمين جميع الأخطار للحفارات البرية والبحرية الخاص بالسوق الإنجليزية، وهو الشائع عالمياً، ويتم تغطية وحدات الحفر أثناء نقلها وإعادة تركيبها من موقع إلي آخر، وأثناء عملية الحفر وأثناء التخزين في حالة عدم التشغيل.

٢/٤ الأخطار المستثناه والتي لا يتم التأمين عليها:

هناك بعض الأخطار التي تكون مستثناه من وثائق تأمين البترول منها ما يلي:

١/٢/٤ لا يغطى التأمين الأجزاء التى تعمل تحت سطح الأرض حيث تستثنى وثائق تأمينات البترول التأمين على الأجزاء التي تعمل بشكل كامل تحت سطح الأرض، وقد تلتصق بعض تلك الأجزاء بالصخور أثناء الحفر دون وقوع حادث انفجار ويصعب استخراجها، وقد تسقط بالبئر نتيجة انفصالها عن ساق الحفر وفي كلا الحالتين لا يتم تعويضها منفردة، ولكن في حالة الانفجار أو الحريق يتم تعويض تلك الأجزاء بالكامل.

٢/٢/٤ تستثنى الوثائق كذلك خطر الإرهاب، والتخريب، والعنف السياسي، والحروب، والاستيلاء علي السلطة والحرب النووية وغيرها.

٣/٤ تغطيات إضافية للمسئولية المدنية قبل الغير:

هناك بعض التغطيات الإضافية فى نشاط البترول والمرتبطة بالحفارات مثل:

١/٣/٤ تأمين المسئولية المدنية العامة، وذلك فى حالة تسبب الحفار في أحداث تلفيات لممتلكات أو أضرار جسمانية تخص الغير – الطرف الثالث – وفى هذه الحالة تلتزم الوثيقة بسداد ما تحكم به المحكمة كتعويض للمتضرر، وقد يصيب الحفار ممتلكات تخص المالك فيتم تعويض المالك عن الأضرار التي تسبب بها الحفار,

وهناك أيضاً تأمين مسئولية رب العمل عن عمال شركة الحفر، فى حالة حدوث حالات وفيات بين العمال، أو أضرار جسمانية أثناء العمل حفر الآبار ويكون المتسبب صاحب العمل, لذا لابد من ثبوت مسئولية صاحب العمل عن الحادث.

٢/٣/٤ فقد الاستعمال الناتج من حادث، فقد يطلب مالك وحدة الحفر تأمين الأرباح التي كان سوف يجنيها من الحفار، لو لم يتوقف بسبب حادث يستلزم الاصلاح وغالباً يتم النص علي مدة محددة يتم خلالها تعويض مالك الحفار بنسبة تعويض تمثل مبلغ تأمين الوثيقة وهي أرباح العقود الفائتة من توقف الحفار بسبب حادث، ومن المعروف أن توقف الحفار نتيجة أعمال الصيانة، أو عدم التعاقد يخرج من نطاق هذه التغطية.

٣/٣/٤ تأمين الأخطار السياسية، وهى وتختلف عن تأمين العنف السياسي ويتم تغطيتها عند تعاقد الحفار علي أنشطة حفر خارج بلده، وتلك الأخطار هي المصادرة، والتأميم، والاستيلاء، والمنع من التصرف، والحجز والتجريد من الملكية.

٥-الدراسة التطبيقية

في هذا الجزء من البحث قام الباحث بتطبيق الخوارزميات الأربعة على فرع تأمين البترول في الشركتين محل البحث، وعرض الأساليب التي استخدمها في تحقيق أهداف البحث وفروضه.

١/٥ منهجية نموذج خوارزمية C4.5

هي سلسلة من الخوارزميات التي يتم استخدامها في مشاكل التصنيف في التعلم الآلي والتنقيب عن البيانات, وتهدف إلى التعلم تحت الإشراف بالنظر إلى مجموعة بيانات معينة، ويمكن وصف كل مجموعة فيها بمجموعة من قيم السمات، وتنتمي كل مجموعة إلى فئة معينة في فئة حصرية متبادلة, وتستخدم خوارزمية C4.5 ما يعرف بمفهوم ربح المعلومة، وذلك من خلال تطبيق مفهوم تقنية التنقيب في البيانات ومفهوم الأنتروبي.

وهذه الخوارزمية قدمت لأول مرة عن طريق العالم الأمريكي **روس كوينلان عام١٩٩٣**باعتبارها امتداد وتطوير لخوارزمية ID3 ، والتي سبق أن قدمها نفس العالم في عام ١٩٨٦, وقد قدمها ليتغلب على العيوب التى كانت في خوارزمية ID3 حيث يحتوي مقياس إكتساب المعلومات الذي تستخدمه ID3على عيب خطير، فهو يفضل عمومًا الميزات التي تحتوي على المزيد من قيم السمات، لأن الميزات التي تحتوي على مزيد من قيم السمات ستحقق مكاسب كبيرة نسبيًا في المعلومات (تعكس مكاسب المعلومات شرطًا معينًا ويجب أن تكون درجة الحد من التيقن هي أنه كلما كانت مجموعة البيانات أدق، كلما زاد التيقن، وكلما صغر حجم الإنتروبيا الشرطية، زاد اكتساب المعلومات) ولتجنب

هذا العيب تم استخدام نسبة كسب المعلومات في .C4.5 كمبدأ توجيهي لاختيار الفروع وتقدم نسبة الحصول على المعلومات مصطلحًا يُسمى تقسيم المعلومات لمعاقبة الميزة الأكثر قيمة ،وبناء على ذلك يمكن القول أن خوارزمية C4.5تتغلبعلى مشكلة عدم تمكن خوارزمية ID3 من التعامل مع قيم سمات الميزة المستمرة، ومع ذلك تحتاج قيم السمات المستمرة إلى المسح والفرز.ويمكن النظر إلى عملية توليد شجرة قرار أولية (initial tree) من مجموعة أمثلة التدريب على أنها أهم جزء في خوارزمية C4.5 ويتم انشاء خوارزمية نموذج تصنيف على شكل شجرة قرار وتتكون بنية شجرة القرار المولدة من نوعين من العقد كما يلي:

- عقدة ورقة،وهي عقدة تعبِّر عن قاعدة نهائية.
- عقدة قرار وهي التي تحدد اختباراً ما على قيمة خاصية مفردة مع فرع وحيد، و شجرة جزئية لكل ناتج محتمل للاختبار.

ومن خلال ذلك يمكن استعمال الشجرة الناتجة لتصنيف مثال جديد بدءاً بعقدة الجذر، مروراً بالمسارات، ووصولاً إلى عقدة ورقة تمثل صفاً ينتمي إليه تلك البيانات.

وتعتمد خوارزمية C4.5 لبناء شجرة قرار من مجموعة أمثلة تدريب T على ثلاث حالات لمحتوى المجموعة T فمثلاً إذا كانت الصفوف المحددة مسبقاً هي {C1, …, Ck} عندئذ يكون:

T_1 تحتوي مثالاً أو أكثر جميعها ينتمي إلى الصف .Cj وشجرة القرار هنا هي ورقة تمثل صنفاً وحيداً.Cj.

T_2 لا تحتوي أمثلة، و شجرة القرار هنا هي أيضاً ورقة و لكن لتحديد الصنف الذي تمثله هذه الورقة نحتاج إلى معلومات جديدة تزودنا بها مجموعة أخرى غير T بمعنى أن T لا تصلح أساساً لتكون مجموعة تدريب.

T_3 تحتوي أمثلة تنتمي إلى مجموعة مختلطة من الأصناف وفي هذه الحالة الفكرة هي "تنقية" أو تجزئة المجموعة T إلى مجموعات جزئية مكونة من أمثلة مع محاولة الاقتراب من تجمع من أمثلة ينتمي إلى صنف وحيد و محدد كلما أمكن ذلك.

ويمكن توضيح خطوات تقنية خوارزمية C4.5 كما يلي:

$$SplitInformation\ (D,A) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|}\ log\frac{|D_i|}{|D|}$$

$$Gain\ Ratio\ (D,A) = \frac{g(D,A)}{SplitInformation\ (D,A)}$$

٢/٥منهجية نموذج خوارزمية CART

إن خوارزمية CART (Classification And Regression Trees) هي نوع من شجرة القرار ذات التقسيم الثنائي الرجعي الذي يقوم بالمراجعة المستمرة للخواص كأهداف وتنبؤات، ويتطور حجم الشجرة إلى أعلى حجم بدون إستخدام قوانين توقف وبالتقسيم المتكرر منالجذر،وتستخدم الخوارزمية التقليم بأسلوب يعرف بأسلوب تعقيد الكلف, وطبقاً لهذا الأسلوب فإن مجموعة متسلسلة من الأشجار تنمو ويتم الحصول على كل واحدة منها من سابقتها باستبدال شجرة ثانوية واحدة أو أكثر مع الأوراق, وتشمل الشجرة الأخيرة في السلسلة ورقة مفردة واحدة, وتلك الورقة هي التي تتنبأ بصنف محدد(Matthew and et al ,2010) Sathyadevi G,2011).

وتعد خوارزمية CART من أفضل عشرة خوارزميات للتعلم الآلي,و تستخدم كشجرة تصنيف فهرس GINI عندما تنقسم العقدة, يستخدم مؤشر GINI بشكل أساسي لقياس شوائب قسم البيانات أو مجموعة بيانات التدريب (D) وكلما كانت قيمة GINI أصغر، زادت درجة نقاء العينة (أي كلما زاد احتمال انتماء العينة إلى نفس الفئة),بعد قياس مؤشر Giniلجميع قيم ميزة معينة في مجموعة البيانات، يمكن الحصول على معلومات Gini Splitالخاصة بالميزة، وهيGiniGain بغض النظر عن التقليم، فإن عملية إنشاء شجرة قرار التصنيف بشكل متكرر هي تحديد أصغر عقدة لـ GiniGain كنقطة تشعب في كل مرة حتى تنتمي مجموعات البيانات الفرعية إلى نفس الفئة أو يتم استخدام جميع الميزات.

وبالنسبة للسمات المنفصلة، قد يكون هناك موقف يكون فيه عدد قيم السمات$N \leq 3$ ، ونظرًا لأن خوارزمية CART عبارة عن شجرة ثنائية،لذا فانه يجب مراعاة وجود فرعين فقط عند معالجة الميزات المنفصلة باستخدام قيم$N \leq 3$ ويجب إنشاء تسلسل ذو قيمتين بشكل مصطنع من خلال الدمج واتخاذ أصغرGiniGain كنقطة قرار تشعب الشجرة,وإذا كانت قيمةGiniGain تحتوي على ثلاث قيم فإن التسلسل الثنائي يحتوي على احتمالات ثنائي(Suknovic, Milija, 2012).

وعند استخدام خوارزمية CART ، من الضروري حساب فهرسGini عند التشعب وفقًا للتسلسل الثنائي، ثم تحديد التسلسل الثنائي الذي ينتج أصغرGINIGain باعتباره التسلسل الثنائي المتشعب لهذه الميزة للمشاركة في تكرار بناء الشجرة.

وقد تم تطبيق خوارزمية CART كما يتضح في المعادلات التالية:
(Sharma &Bhartiya, 2012) ,(Suknovic, Milija, 2012)

$$G(x, S) = I(S) - PL \times (SL) - PR \times I(SR)$$

$$I(S) = \sum_{i=j} (P\,(I/S)P\,(j/S)\,, I \neq j$$

احتمال وجود المتغيرات في العقدة P(i|S)

احتمال الحالات المرسلة إلى الفرع الأيمن من شجرة القرار (PR)
العقد اليمنى الناتجة من الجيل الجديد (الأبناء). (SR)
العقد اليسرى الناتجة من الجيل الجديد (الأبناء). (SL)

$$Gini\ (D) = 1 - \sum_{i=0}^{n} \left(\frac{D_i}{D}\right)^2$$

$$Gini\ \left(\frac{D}{A}\right) = \sum_{i=0}^{n} \frac{D_i}{D}\ Gini(D_i)$$

٣/٥ منهجية نموذج خوارزمية LogisticRegression

تعتبر خوارزمية Logistic Regression أحد خوارزميات التصنيف، والتي تُستخدم عندما يكون متغير الاستجابة فئوي, وتقوم فكرة الانحداراللوجستي على إيجاد علاقة بين الخَصائص واحتمال

نتيجة معينة, وهي تقوم بتصنيف البيانات إلى فئات منفصلة, ويتم ذلك عن طريق تحويل المخرجات إلى قيم إحتمالية وتصنيفها إلى فئات مختلفة ضمن الخوارزمية, وتم استخدام خوارزمية Logistic– Regression الثنائية في هذا البحث, وتم التنبؤ للبيانات المستقبلية عن طريق المعادلة التالية:

$$h(x_i) = \theta^t x_i$$

حيث $h(x_i)$ هي الفرضية الأساسية, θ هي معامل الانحدار , x_i قيم المتغيرات الفعلية, ولكي تكون المخرجات في شكل احتمال، وتكون القيم محصورة بين $(0,1)$ ، لذا سيتم تطبيق دالة Sigmoid من خلال الصيغة التالية :

$$h(x_i) = g(\theta^T x_i) = \frac{1}{1 + e^{-\theta^T x_i}}$$

ويكون تعريف الاحتمالية في النموذج، والذي يعطي احتمالية التصنيف للفئتين $(0,1)$ كما يلي:

$$\boldsymbol{P(y_{i=1} I x_i \theta) = h(x_i)}$$
$$\boldsymbol{P(y_{i=0} I x_i \theta) = 1 - h(x_i)}$$

وبدمج المعادلتين السابقتين نحصل على الصيغة التالية:

$$P(y_i I x_i \theta) = h(x_i)^{y_i}\{1 - h(x_i)^{1-y_i}\}$$

وبعد الحصول على قيمة الاحتمال نوجد دالة الخطأ ((Error function)، وتأخذ الصيغة الرياضية التالية $\iota B = \sum_{i=1}^{n} -y_i \, log h(x_i) - (1 - y_i) \log\{(1 - h(x_i)\}$

٤/٥ منهجية نموذج خوارزمية NaÏve Bayes

تعتبر هذه التقنية من أهم تقنيات التصنيف، وتعتمدعل بنظرية بايز مع افتراض الاستقلال بين المتغيرات, إذ يفترض مصنف NaÏve Bayes أن وجود خاصية أو سمة معينة في فئة لاعلاقة لها بوجود أي خاصية أو سمة أخرى،أي أن مصنف NaÏveBayes سوف يأخذ في الاعتبار كل

الخصائص للمساهمة بشكل مستقل في تحديد المصنف,ويتميز هذا النموذج بالبساطة وسهولة البناء والتعامل مع مهام متعددة التصنيفات, وبكونه مفيد بشكل خاص لمجموعات البيانات الكبيرة (Big data), وأن له كفاءة تصنيفية ثابتة لعدم حساسيته للبيانات المفقودة , وإن كان يعاب على تلك الخوارزمية أنه على الرغم من إحتوائها من الناحية النظرية على أصغر معدل خطأ مقارنة بطرق التصنيف الأخرى إلا أن هذا ليس هو الحال دائمًا، وذلك لافتراض هذا النموذج أن السمات مستقلة عن بعضها البعض في حالة فئة مخرجات معينة, وغالبًا ما يكون هذا الافتراض غير صحيح في التطبيقات العملية حيث أن هناك المزيد من السمات عندما يكون الارتباط كبير فإن تأثير التصنيف ليس جيدًا، وعندما يكون ارتباط السمة صغير يكون أداء خوارزمية Naïve Bayes أفضل،وقد تم إدخال تحسينات على هذه الخوارزمية من خلال النظر في الصلة الجزئية, ومن الضروري معرفة الاحتمال السابق، ويعتمد الاحتمال السابق غالبًا على الفرضية، لذا يمكن أن يكون هناك العديد من أنواع النماذج الافتراضية في بعض الحالات ويكون ذلك بسبب النموذج الافتراضي السابق، مما قد يؤدي إلى سوء التنبؤ, وقد تم التغلب على هذه المشكلات في البحث الحالي حيث تم إدخال العديد من السمات المؤثرة في مطالبات فرع تأمين البترول، ودراسة مدى تأثير كل سمة من السمات على المطالبات, كما تم استخدام صيغة الاحتمال الكلي، وصيغة الاحتمال العكسي طبقاً لصيغة الإحتمال الشرطي، وحل الإحتمال اللاحق وتصنيفه طبقاً للإحتمال السابق, وذلك من خلال معادلة الإحتمال الشرطي الآتية:

$$P(Y = y_r|X = x_1) = \frac{P(X = x_1|\ Y = y_r)p(y = y_r)}{P(X = x_r)}$$

وحيث أن $P_q + P_r = 1$

$$P(Y = y_q|X = x_1) = \frac{P(X = x_1|\ Y = y_q)p(y = y_q)}{P(X = x_q)}$$

ويكون لمجموعة العينات n مجموعة من المميزات i=1,2,3,….,m وندخل مجموعة من التدريبات على النحو التالي:

$$D = \{(x_1^{(1)}, x_2^{(1)}, \dots x_n^{(1)}, \Upsilon_1), (x_1^{(2)}, x_2^{(2)}, \dots x_n^{(2)}, \Upsilon_2) \dots (x_1^{(m)}, x_2^{(m)}, \dots x_n^{(m)}, \Upsilon_m)\}$$

والقيم { C_1,C_2,….C_kC1,C2,C3,…ck }

الإحتمال الأول $P(y=C_k$,k=1,2,3,.....k)

وتكون صيغة الاحتمال بناء على افتراض الاستقلال المشروط

$$P(X = x|Y = C_k) = P(X_1 = x_1, X_2 = x_2, \dots \dots X_n|Y = C_k)$$

$$C_k \prod_{j=1}^{n} P\ (X_j = x_j|Y = C_k)$$

ويأخذ الإحتمال العكسي الصيغة التالية:

$$P(Y = C_k|X = x) = \frac{P\ (X = x|Y = C_k)P(Y = C_k)}{\sum_k (X = x|Y = C_k)P(Y = C_k)}$$

ويتم تقدير معلمات نموذج خوارزمية Naïve Bayes كما يلي :

$$\boldsymbol{P\ (Y = C_k = \frac{\sum_{i=1}^{m} \mathbf{I} x_j^i = \ a_{j1,y1} = C_k}{\sum_{i=1}^{m} \mathbf{I}(y_1 = C_k)}}$$

٥/٥ معالجة البيانات ومتطلبات التشغيل

هناك فرق بين تقنيات التنقيب في البيانات وأدوات التنقيب في البيانات (Aksenova S Svetlana,2004) , ففي حين يشير مصطلح تقنيات التنقيب في البيانات إلى التقنيات المختلفة المستخدمة في هذا المجال والتي سبق التنويه عنها فإن أدوات التنقيب تعبر عن تلك البرامج التي تسمح للمستخدمين استخدامها في تحليل البيانات بطرق وحزم برامج جاهزة, ومن أشهر برامج التنقيب في البيانات برنامج WEKA (Waikato Environment for Knowledge Analysis) وهو برنامج مجاني ومفتوح المصدر ويعتمد في استعمالة على JAVA ويمكن استخدامه من خلال Windows , Mac OSX, Linux ويتميز بأن له العديد من الوظائف مثل معالجة البيانات، واختيار الميزات، والتصنيف، هذا بالإضافة إلى تقديم وظيفة التجميع، وقواعد الارتباط، والتصوّر، وغيرها من الوظائف الأخرى, وهناك برامج أخرى للتنقيب في البيانات مثل برنامج Raper Miner

وبرنامج DataMelt أو DMelt ويتميز هذا البرنامج بأنه يمكن استخدامه من خلال Windows Mac OSX, Linux , بالإضافة إلى Android لذا فهو يفضل لإستخدامات المهندسين والحالات العملية للطلاب, وهناك برنامج (Knowledge Extraction for Evolutionary Learning KEEL) وهذا البرنامج يسمح بتنسيق ملفات مختلفة ويقوم بالمعالجة المسبقة للبيانات وهو يستند أيضاً إلى JAVA ويستخدم في الأغراض البحثية والتعليمية, وبرنامج Massive Online Analysis (MOA) وهو يستند إلى JAVA ويفضل استخدامه عندما يكون المطلوب استخدام التطبيقات في الوقت الفعلي ويمكن ربطه ببرنامج WEKA .

وفي هذا البحث قد تم الإعتماد على البيانات الربع سنوية للعينة المختارة **(شركة مصر للتأمين وشركة قناة السويس للتأمين)** خلال الفترة من ٢٠١١ حتى ٢٠٢٠، طبقاً للسمات الآتية:صافي الأقساط، مخصص الأخطار السارية أول المدة، مخصص الأخطار السارية آخر المدة، الأقساط المكتسبة، صافي التعويضات, مخصص التعويضات تحت التسوية أول المدة, مخصص التعويضات تحت التسوية آخر المدة, التعويضات التحميلية,فائض/عجز النشاط التأميني, وعدد المطالبات, كما تم معالجة وتشغيل البيانات بإستخدام برنامج (WEKA), وقد قام الباحث بتشغيل هذا البرنامج على حاسب شخصي (Personal Computer) مكوناته الرئيسية هي: Processor Intel Core i5 ، 16 GB RAM ، 500GB Hard Disk

٦/٥ أهم معلمات النموذج :

لتقييم دقة نماذج التنبؤ تم استخدام ثلاثة معلمات إحصائية ، وهي معامل الارتباط (r)، ومتوسط الخطأ التربيعي (RMSE)، وقيمة t (إحصاءة t), ويمكن عرض كل معلمة من المعلمات الإحصائية الثلاثة على النحو التالي:

١/٤/٥ معامل الارتباط Correlation coefficient(r)

ويحدد معامل الارتباط العلاقة الخطية بين متغيرات الإدخال (x) والمتغيرات المستهدفة (y), وهومعامل تكون قيمته بين (−١ و ١) ويتم الحصول عليه من خلال العلاقة التالية:

$$r = \sqrt{\left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})(P_i - \bar{P})}{n\, S_y S_p}\right]^2}$$

حيث

n: عدد المفردات.

y: المخرجات الفعلية.

p: القيم المتوقعة.

$\overline{P}$ و $\overline{y}$:متوسطات المخرجات الفعلية والقيم المتوقعة على التوالي.

و S_y و S_p انحرافات معيارية للمخرجات الفعلية والقيم المتوقعة على التوالي.

٢/٤/٥ الجزر التربيعي لمربع متوسط الخطأ Root Mean Square Error (RMSE)

وهو عبارة عن الجذر التربيعي الحسابي لمتوسط الخطأ التربيعي، ويتم حسابه من خلال المعادلة التالية:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - y_i)^2}$$

حيث :

n:عدد مشاهدات المخرجات.

p:القيم المتوقعة.

y:ملاحظات المخرجات الفعلية.

٣/٤/٥ قيمة t (إحصاءة t).

بالنظر إلى قيمة معامل ارتباط بيرسون r، وملاحظات الإخراج للحجم n ، فإن القيمة المعيارية t لمعامل ارتباط بيرسون (مهمة عند $\alpha = 0.05$) هي:

$$t = \sqrt{\frac{r}{\frac{1-r^2}{n-2}}}$$

حيث :

$Df = n-2$

ونظرًا لوجود df واحد لكل عمودفي حين أن معامل الارتباط ليس قريبًا جدًا من ١ ، إلا أنه يكفي أكبر من ٠ لرفض فرضية عدم وجود علاقة بين المتغيرات وقبول الفرضية البديلة القائلة بوجود علاقة ذات دلالة إحصائية بين المتغيرات.

٧/٥إختبار الفروض الإحصائية

تمكن الباحث بعد تطبيق الخوارزميات الأربعة سابقة الذكر على فرع تأمين البترول في شركتي مصر للتأمين وقناة السويس للتأمين من الوصول إلى النتائج الموضحة بالجدول التالي، وذلك عن طريق استخدام واجهة برنامج WEKA،والتي أعطت ٦ معايير إحصائية هي معامل التحديد R^2,اختبار t-tes , متوسط الخطأ المطلقMAE , و الجذر التربيعى لمتوسط مربعات الخطأ RMSE ,و الجذر التربيعى لمتوسط الخطأ المطلق النسبي RRSE , متوسط الخطأ المطلق النسبي MRAE

جدول رقم (١/٥)

المقارنة بين نتائج تطبيق استخدام خوارزميات –NaÏve Bayes ,Logistic Regres sion ,CART , C4.5في النمذجة الاكتوارية للمطالبات بفرع تأمين البترول بشركتي مصر للتأمين وقناة السويس للتأمين

المعايير الإحصائية	شركة مصر للتأمين				شركة قناة السويس للتأمين			
	خوارزمية C4.5 (J48)	خوارزمية CART (REPTree)	خوارزمية Logistic Regres–sion	*خوارزمية* NaÏve Bayes	خوارزمية C4.5 (J48)	خوارزمية CART (REPTree)	خوارزمية Logistic Regres–sion	خوارزمية NaÏve Bayes
١– معامل التحديد R^2	78.6 %	88.1 %	85%	**90%**	81.3 %	89.1 %	85.3 %	**93.1%**
٢– اختبار *t*–test	0.46	0.59	0.50	**0.63**	0.38	0.46	0.42	**0.53**
٣– متوسط الخطأ المطلقMAE Mean Absolute error	0.139	0.113	0.125	**0.111**	0.158	0.141	0.143	**0.122**
٤– الجذر التربيعى لمتوسط مربعات الخطأ RMSE	0.3935	0.1322	0.1575	**0.1143**	0.3256	0.2356	0.2584	**0.1053**

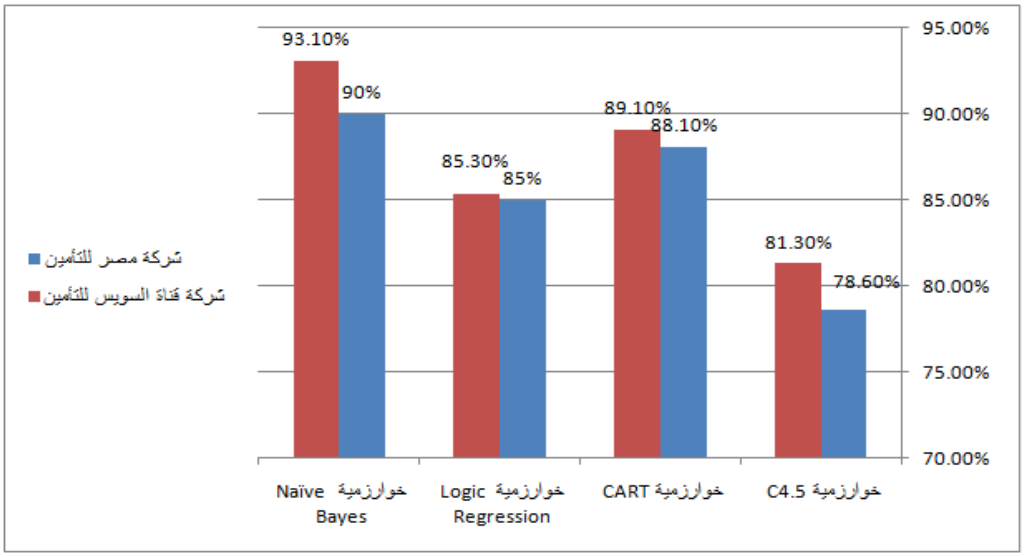
								Root Mean Squared error
30.146 1%	34.1235 %	33.2433 %	34.1552 %	**30.112 7%**	35.2325 %	34.9213 %	40.1256 %	٥- متوسط الخطأ المطلق النسبى MRAE Mean Relative absolute error
67.105 3%	73.2234 %	75.2356 %	80.1256 %	**60.124 3%**	80.1355 %	76.3422 %	82.1345 %	٦- الجذر التربيعى لمتوسط الخطأ المطلق النسبيRRSE Root Relative Squared Error

من تحليل بيانات جدول (١/٥) والمتعلق بالمقارنة بنتائج تطبيق الخوارزميات الأربعة في النمذجة الاكتوارية للمطالبات بفرع تأمين البترول بشركة مصر للتأمين وقناة السويس للتأميناً أمكن الحصول على التحليلات التالية:

١/٧/٥ مقارنة القدرة التفسيرية للنموذج من خلال معامل التحديد (R^2) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين

شكل ١/٥

معامل التحديد (R^2) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين



من خلال تحليل نتائج جدول (١/٥)، والشكل (١/٥) تم التوصل للنتائج التالية :

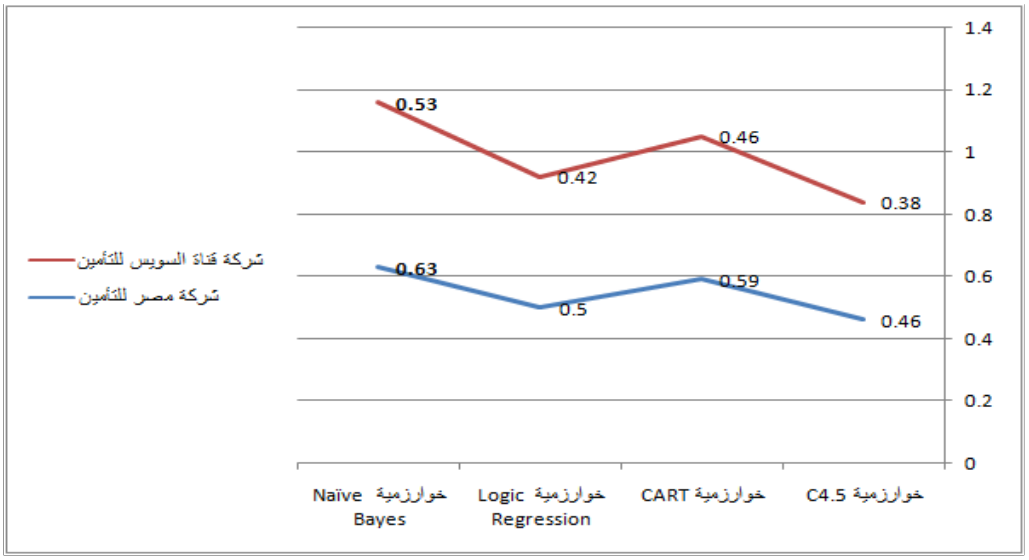
أن المتغيرات الخارجية تعمل على تفسير التغير في مطالبات فرع تأمين البترول في شركة مصر للتأمين بنسبة (٧٨,٦٪) مقابل نسبة (٨١,٣%) في شركة قناة السويس للتأمين عند استخدام خوارزمية C4.5 ، و على نسبة (٨٨,١%) في شركة مصر للتأمين مقابل نسبة (٨٩,١%) في شركة قناة السويس للتأمين عند استخدام خوارزمية CART ، وكذلك على نسبة (٨٥%) في شركة مصر للتأمين مقابل نسبة (٨٥,٣٪) في شركة قناة السويس للتأمين عند استخدام خوارزمية Logistic Regression ، وأخيراً نسبة (٩٠٪) في شركة مصر للتأمين مقابل نسبة (٩٣,١٪) في شركة قناة السويس للتأمين عند استخدام خوارزمية NaÏve Bayes، وذلك فيما يتعلق بالقدرة التفسيرية للنموذج من خلال معامل التحديد (R^2), وبالنظر للقيم المختلفة لمعامل التحديد (r^2) تبين أن أعلى قيمة له كانت عند تطبيق خوارزمية NaÏve Bayes مما يعطي مؤشراً بأن المتغيرات الخارجية عند تطبيق هذه الخوارزمية تعمل على

تفسير مطالبات فرع تأمين البترول في الشركتين محل الدراسة أكثر من تطبيق باقي الخوارزميات حيث كانت النسبة في الشركتين ٩٠٪ , ٩٣,١٪.

٢/٧/٥ مقارنة جودة توفيق النموذج من خلال اختبار (t- test) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين

شكل (٢/٥)

اختبار (t- test) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين

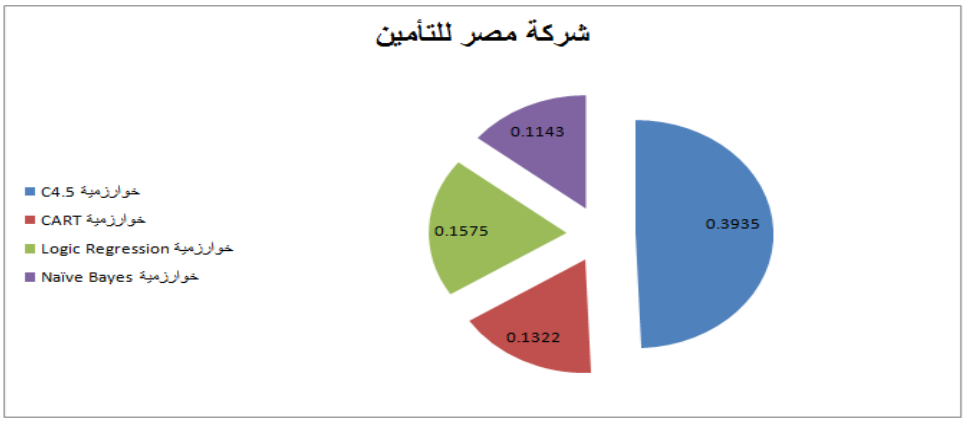


تبين من تحليل جدول (١/٥) والشكل (٢/٥) أنه فيما يتعلق بجودة توفيق النموذج من خلال إختبار t-Test ، نجد قيمة (t) قد بلغت (٠,٩٦), (٠,٣٨) عند استخدام خوارزمية C4.5 في شركة مصر للتأمين وشركة قناة السويس للتأمين على الترتيب وهي دالة إحصائياً, وعند استخدام خوارزمية CARTبلغت (٠,٥٩) , (٠,٤٦) في الشركتين محل الدراسة على الترتيب, بينما بلغت (٠,٥٠), (٠,٤٢) عند استخدام خوارزميةLogistic Regression ، وأخيراً بلغت (٠,٦٣) , (٠,٥٣) عند استخدام خوارزمية NaÏve Bayes وهي أيضاً دالة إحصائياً, ويمكن من خلال النتائج المتعلقة بجودة توفيق النموذج، القول بأن تطبيق خوارزمية NaÏve Bayes أعطت أفضل نتيجة في الشركتين محل الدراسة.

٣/٧/٥ مقارنة معايير أخطاء النموذج باستخدام مقياس الجذر التربيعى لمتوسط مربعات الخطأ RMSEلكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين

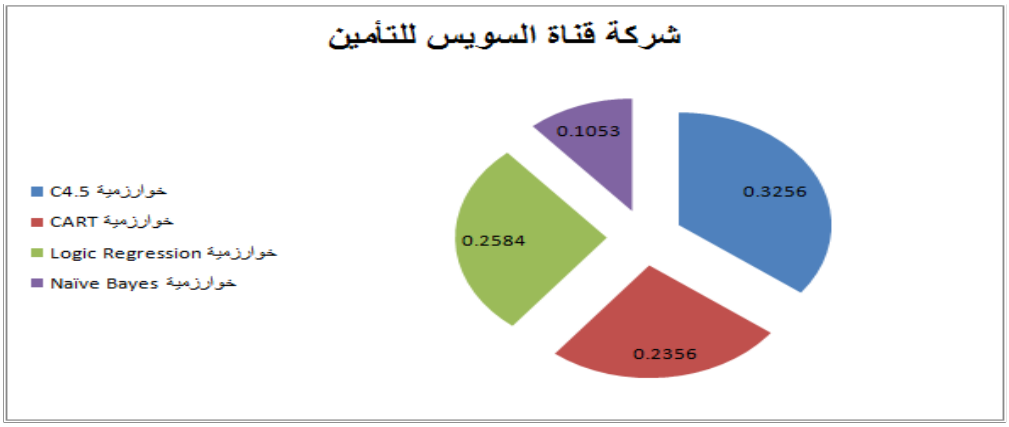
شكل (٣/٥)

مقياس RMSE لكل من الخوارزميات الأربعة بالتطبيق على مصر للتأمين



شكل (٤/٥)

مقياس RMSE لكل من الخوارزميات الأربعة بالتطبيق على قناة السويس للتأمين



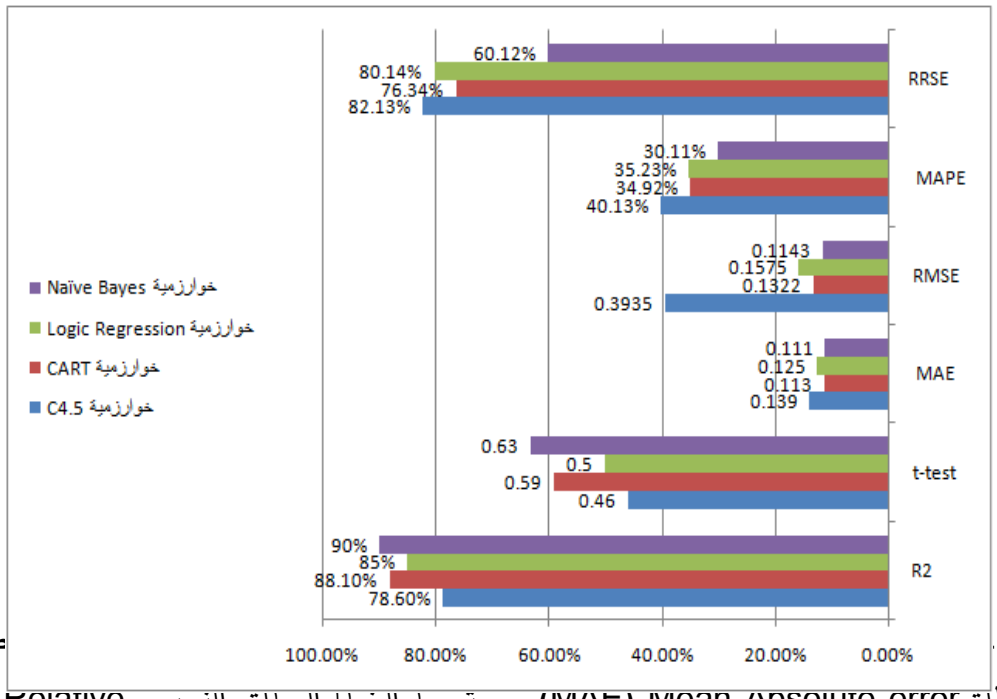
من تحليل نتائج جدول (١/٥) والشكلين (٣/٥) , (٤/٥) تبين أن قيمة الجذر التربيعى لمتوسط مربعات الخطأ RMSEقد بلغت (٠,٣٩٣٥), (٠,٣٢٥٦) عند استخدام خوارزمية C4.5 في شركة

مصر للتأمين وشركة قناة السويس للتأمين على الترتيب, وعند استخدام خوارزمية CART بلغت (٠,١٣٢٢) , (٠, ٢٣٥٦) في الشركتين محل الدراسة على الترتيب, بينما بلغت (٠,١٥٧٥) ,(٠,٢٥٨٤) عند استخدام خوارزميةLogistic Regression ، وأخيراً بلغت (٠,١١٤٣) , (٠,١٠٥٣) عند استخدام خوارزمية Naive Bayes, ويتضح أن أقل قيمة لمتوسط مربعات الخطأ RMSEعند تطبيق خوارزمية Logistic Regression مما يعطي مؤشراً أن أخطاء تطبيق هذا النموذج أقل من باقي النماذج, وبالتالي تفوق خوارزمية Naïve Bayes من حيث أخطاء النموذج على باقي الخوارزميات في كلاً من الشركتين محل الدراسة.

٤/٧/٥ المقارنة بين مؤشرات متوسط الخطأ المطلقMean Absolute error (MAE) , متوسط الخطأ المطلق النسبى Mean Relative absolute error(M RAE) ، والجذر التربيعى لمتوسط الخطأ المطلق النسبيRoot Relative Squared Error (RRSE) عند تطبيق الخوارزميات الأربعة في شركة مصر للتأمين

شكل (٥/٥)

المقارنة بين مؤشرات الإختبار في شركة مصر للتأمين



من ط الخطأ

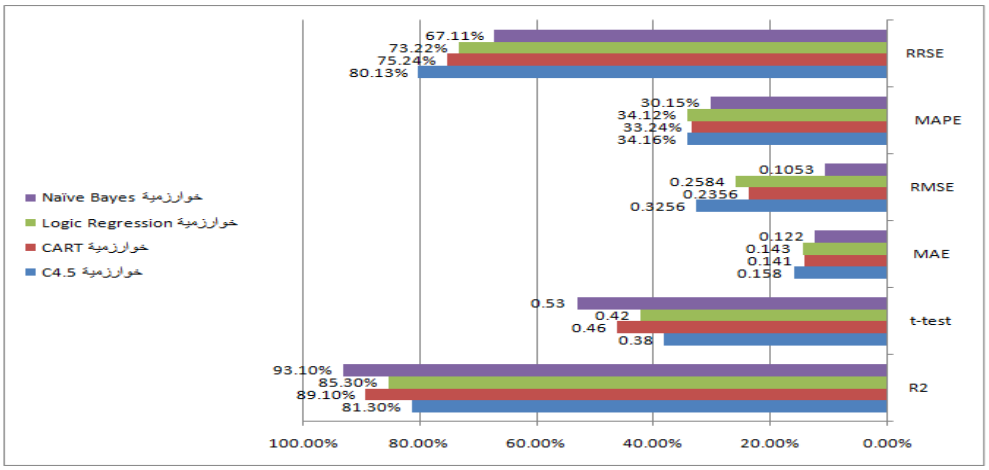
المطلقMean Absolute error (MAE) , متوسط الخطأ المطلق النسبى Mean Relative

(RRSE) Root والجذر التربيعى لمتوسط الخطأ المطلق النسبي absolute error (MRAE) Relative Squared Error عند تطبيق الخوارزميات الأربعة في شركة مصر للتأمين, حيث إن قيم متوسط الخطأ المطلق (MAE) عند تطبيق الخوارزميات الأربعة على الترتيب كانت ٠,١٣٩ , ٠,١١٣, ٠,١٢٥ , ٠,١١١ ، ومن الواضح أن أقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes, وكانت قيم متوسط الخطأ المطلق النسبىMRAE عند تطبيق الخوارزميات الأربعة على الترتيب ٤٠,١٢٥٦ , ٣٤,٩٢١٣ , ٣٥,٢٣٢٥ , ٣٠,١١٢٧ ، ومن الواضح أيضاً أن أقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes , وأخيراً كانت قيم مؤشر الجذر التربيعي لمتوسط الخطأ المطلق النسبي RRSE كانت ٨٢,١٣٤٥ , ٧٦,٣٤٢, ٨٠,١٣٣٥٣ , ٦٠,١٢٤٣ على الترتيب وأقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes, ومما سبق يمكن القول أنه قد أمكن تطبيق الخوارزميات الأربعة على فرع البترول في شركة مصر للتأمين، وأن تطبيق خوارزمية NaÏve Bayes قد تفوقت على باقي الخوارزميات الأخرى عند مقارنة جميع المؤشرات.

٥/٧/٥ المقارنة بين مؤشرات متوسط الخطأ المطلقMean Absolute error (MAE) , متوسط الخطأ المطلق النسبى Mean Relative absolute error(MRAE)، والجذر التربيعى لمتوسط الخطأ المطلق النسبيRoot Relative Squared Error (RRSE) عند تطبيق الخوارزميات الأربعة في شركة المهندس للتأمين

شكل (٦/٥)

المقارنة بين مؤشرات الإختبار لشركة قناة السويس للتأمين



من خلال تحليل بيانات الجدول (١/٥) والشكل (٦/٥) يمكن المقارنة بين مؤشرات متوسط الخطأ المطلقMean Absolute error (MAE) , متوسط الخطأ المطلق النسبى Mean Relative absolute error (MRAE) والجذر التربيعى لمتوسط الخطأ المطلق النسبي Root (RRSE) Relative Squared Error عند تطبيق الخوارزميات الأربعة في شركة قناة السويس للتأمين, حيث إن قيم متوسط الخطأ المطلق (MAE) عند تطبيق الخوارزميات الأربعة على الترتيب كانت ٠,١٥٨ , ٠,١٤١, ٠,١٤٣ , ٠,١٢٢ ، وأن أقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes, وكانت قيم متوسط الخطأ المطلق النسبى MRAE عند تطبيق الخوارزميات الأربعة على الترتيب ٣٤,١٥٥٢ , ٣٣,٢٤٣٣ , ٣٤,١٢٣٥ , ٣٠,١٤٦١، ومن الواضح أيضاً أن أقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes , وأخيراً كانت قيم مؤشر الجذر التربيعي لمتوسط الخطأ المطلق النسبي RRSE كانت ٨٠,١٢٥٦ , ٧٥,٢٣٥٦, ٧٣,٢٢٣٤ , ٦٧,١٠٥٣ على الترتيب، وأقل قيمة لها كانت عند تطبيق خوارزمية NaÏve Bayes , ومما سبق يمكن القول أنه قد أمكن تطبيق الخوارزميات الأربعة على فرع البترول في شركة قناة السويس للتأمين، وأن تطبيق خوارزمية NaÏve Bayes قد تفوقت على باقي الخوارزميات الأخرى عند مقارنة جميع المؤشرات.

٦- النتائج والتوصيات

١/٦ النتائج

في إطار هدف ومنهج البحث،وفي ظل فروضه توصل الباحث إلى مجموعة من النتائج أهمها:

- أنه أمكن تطبيق الخوارزميات الأربعة على فرع تأمين البترول في كل من شركة مصر للتأمين وشركة قناة السويس للتأمين, وهذا يؤكد صحة الفرض البحثي الأول والقائل بأنه"**يمكن استخدام خوارزميات مختلفة للتنقيب في البيانات في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول**".
- أنه من الدراسة التطبيقية تبين أن هناك إختلاف بين تطبيق النماذج المختلفة للخوارزميات المستخدمة في البحث في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول, وأن خوارزميةNaÏve Bayes أعطت أفضل النتائج مما يدل على عدم صحة الفرض الثاني

والقائل بأنه "**لا يوجد إختلاف بين تطبيق النماذج المختلفة للخوارزميات المستخدمة في البحث في النمذجة الاكتوارية للمطالبات في التأمينات العامة فرع تأمين البترول**".

- عند مقارنة القدرة التفسيرية من خلال معامل التحديد (R^2) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين؛ تبين أن أعلى قيمة لمعامل التحديد كانت عند تطبيق خوارزمية Naïve Bayes مما يعطي مؤشراً بأن المتغيرات الخارجية عند تطبيق هذه الخوارزمية تعمل على تفسير مطالبات فرع تأمين البترول في الشركتين محل الدراسة أكثر من تطبيق باقي الخوارزميات.
- عند مقارنة جودة توفيق النموذج من خلال اختبار (t test) لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين؛ تبين أن استخدام خوارزمية Naïve Bayes دالة إحصائياً وأنها أعطت أفضل نتيجة في الشركتين محل الدراسة.
- عند مقارنة معايير أخطاء النموذج باستخدام مقياس الجذر التربيعى لمتوسط مربعات الخطأ RMSE لكل من الخوارزميات الأربعة بالتطبيق على شركتي مصر للتأمين وقناة السويس للتأمين؛ تبين تفوقخوارزمية Naïve Bayes من حيث أخطاء النموذج على باقي الخوارزميات في كل من الشركتين محل الدراسة.
- تبين عند المقارنة بين مؤشرات متوسط الخطأ المطلقMean Absolute error (MAE) , متوسط الخطأ المطلق النسبى Mean Relative absolute error (MRAE)، والجذر التربيعى لمتوسط الخطأ المطلق النسبي Root Relative Squared Error (RRSE) عند تطبيق الخوارزميات الأربعة في شركتي مصر للتأمين وقناة السويس للتأمين، أن تطبيق خوارزمية Naïve Bayes قد تفوقت على باقي الخوارزميات الأخرى عند مقارنة جميع المؤشرات.

٦/٢التوصيات

في ضوء الدراسة التطبيقية، ونتائج البحث توصل الباحث لمجموعة من التوصيات هي:

- ضرورة إعطاء إدارات شركات التأمين الاهتمام الكافي الذي يستحقه مفهوم تقنية تنقيب البيانات والخوارزميات العاملة عليها للوصول إلى الأهداف التي تسعى إلبها.
- ضروة استفادة شركات التأمين المصرية من البيانات الضخمة المتاحة لديها .

- ضرورة استخدام شركات التأمين للنماذج الاكتوارية في تحديد سلوك المطالبات تمهيداً لتحديد السعر المناسب.
- اعتماد كل من (خوارزميةC4.5 و خوارزمية CART وخوارزمية Logistic Regressionوخوارزمية NaÏve Bayes) في نمذجة المطالبات في فرع تأمين البترول حيث إنهم من أفضل خوارزميات تنقيب البيانات لنمذجة المطالبات في التأمينات العامة.
- ضرورة تطبيق خوارزمية NaÏve Bayes فيما يتعلق بنمذجة المطالبات في فرع تأمين البترول لأنها أعطت أفضل النتائج.
- ضرورة تطبيق الخوارزميات الأخرى المختلفة في مجال نمذجة المطالبات في جميع فروع التأمين في شركات التأمين.

٧- المراجع

١/٧ المراجع العربية

- إبراهيم, أحمد إســـماعيل محمد, وآخرون, **اســـتخدام تنقيب البيانات في التنبؤ بحجم المستفيدين من الصندوق القومي للتأمين الصحي "المستوى الإقليمي"** رسالة ماجستير غير منشــورة, كلية علوم الحاســوب وتقنية المعلومات, جامعة النيلين,الخرطوم, الســودان, ٢٠٢٠.
- البكري, محمد عبدالســـلام عبدالعزيز,**نموذج مقترح لتطوير عملية المراجعة باســـتخدام أســـلوب التنقيب في البيانات**, المجلة العلمية لقطاع كليات التجارة, جامعة الأزهر, كلية التجارة, العدد ١٦, ٢٠١٦.
- رفاعي, رهف حريري, **إســـتخدام تقنيات التنقيب عن البيانات في دعم قرارات شـــركات التأمين**, رسالة ماجستير غي منشورة, كلية الاقتصاد, جامعة حلب, ٢٠١٧.
- زنرني, هبه محمد أمير وآخرون, **بناء شـــجرة القرار باســـتخدام خوارزمية C4.5 لدعم قرارات الجمعيات الخيرية: دراســـة تطبيقية على جمعية أهل الخير بحلب**, مجلة جامعة القدس المفتوحة للبحوث الإدارية والاقتصـــادية,جامعة القدس المفتوحة, مجلد ٥, عدد ١٣, ٢٠٢٠.
- ســـيد, أحمد فايز أحمد, "**أدوات التنقيب عن البيانات مفتوحة المصـــدر دراســـة تحليلية وتقييمية**" , مجلة جامعة طيبة للآداب والعلوم الإنسانية, ع ١٠ , السنة الخامسة, ٢٠١٦.
- سليمان, أسامة ربيع أمين, **بناء شجرة قرارات تصنيفية كأداة تنقيب عن البيانات Data Mining للتنبؤ بمستوى الاحتفاظ بالأقساط في سوق التأمين المصري باستخدام خوارزمية C4.5** , المجلة العلمية للبحوث التجارية, كلية التجارة , جامعة المنوفية, عدد ١, يناير ٢٠١٩.
- علي, هدى عبد الرحيم حســين, **اســتخدام تقنية التنقيب عن البيانات لتحليل المؤشــرات المالية لعينة من المصـــارف الأهلية العراقية باعتماد خوارزمية CART** ,المجلةالعراقية لتكنولوجيا المعلومات,الجمعية العراقية لتكنولوجيا المعلومات, المجلد التاســـع , عدد ٢, ٢٠١٨.

- عطا, محمد محمد محمد," **تقييم النشاط التأميني المباشر لقطاع البترول المصري والتنبؤ به بالتطبيق على خطر الحريق والانفجار**", المجلة العربية للإدارة, المنظمة العربية للتنمية الإدارية, المجلد ٣٠, عدد ٢, ديسمبر ٢٠١٠م.
- مؤمن , شريف محمد لطفي, **التعدين المالى للبيانات لدعم الممارسات الرقابية بهدف رفع كفاءة النظم المحاسبية الرقمية**,مجلة الفكر المحاسبي, كلية التجارة, جامعة عين شمس, ٢٠١٩.
- نائب, إبراهيم عبدالواحد, **استخدام أشجار القرار لتقييم المشاريع الإنشائية في المنظمات الحكومية: (دراســـة تطبيقية على مديرية الخدمات الفنية بحلب)**, مجلة جامعة القدس المفتوحة للبحوث الإدارية والاقتصادية, جامعة القدس المفتوحة, عدد ١١, ٢٠١٩.

٢/٧ المراجع الأجنبية

- Quinlan, J. R. **"C4.5 Programs for Machine Learning , Morgan Kaufmann Publishers,**" 1993.
- Doskocil R,2016- **An Evaluation oftotal project risk based on fuzzy logic**. 2015..
- Bhowmik R, **"Detecting Auto Insurance Fraud by Data Mining Techniques**." , Journal of Emerging Trends in Computing and Information Sciences،2(4,2011.
- Akinsola Adeniyi F., Sokunbi M.A., Lawal.O.N., Okikiola F.M-, **" A Data Mining Approach To Insurance Risk Analysis. "** International Journal of Engineering and Computer Science ,2015.
- Aksenova S Svetlana,**Macheine learning with WEKA Explorer tutrorial for WEKA , Computer science** ,California State University Sacramento California, 2004.
- Atre,Shaku **Defining Today,S Data Mining, Executive Update Business Intelligence Advisory Service** , CUTTER , Vol.1,2001.
- Matthew Anyanwu N. and Sajjan G. Shiva, ' **Comparative Analysis of Serial Decision Tree Classification Algorithms'**, International Journal of Computer Science and Security, (IJCSS) Volume (3): Issue (3),USA, 2010.
- Sathyadevi G),**Application of Cart Algorithm in Hepatitis Diagnosis**, IEEE –International Conference on Recent Trends in

Information Technology, ICRTIT, MIT, Anna University, Chennai, India,2011.

- Karloff,H**,"Linear Programming",Birkhauser**,Boston,Berlin,22009.
- Vaidehi,R,Predictive **modeling to improve Succision Rate of Bank Direct Marketing Campaign**,International Journal of Management &Business Studies,2016.
- Han,Jand Kamber,M, **Data Mining: Concepts and Techniques** ,(Morgan Kaufman Series of data management systems),San Diego: Academic press,2001.
- Nisbet, and et al, **Handbook of Statistical Analysis & Data Mining Applications**, Academic Press/Elsevier,2009.
- Jože , **Fuzzy Decision Trees as Decision – Making Framework in the Public Sector**,, Yugoslav Journal ofOperations Research, 2011.
- Esraa H. Abd Al-Ameer, **English Text Classification Using Improved Recursive Feature Elimination(IRFE) Algorithm**, Journal of Engineering Sciencesand Information Technology ,Volume (4), Issue (2) 2020.
- Matthew Anyanwu N. and Sajjan G. Shiva ,' **Comparative Analysisof Serial Decision Tree Classification Algorithms'**, International Journalof Computer Science and Security, (IJCSS) Volume (3): Issue (3),USA,2010.
- Suknovic, Miliji and et al , "**Reusable Components In Decision Tree Induction Algorithms"**, ComputStat ,27,2012.
- Werner, M, **"Financial Process mining – Accounting Data Structure Dependent Control Flow Inference"**, International Journal of Accounting Information Systems, Issue 2017.

ملاحق البحث

ملحق (١)

السنة	٢٠١١	٢٠١٢	٢٠١٣	٢٠١٤	٢٠١٥	٢٠١٦	٢٠١٧	٢٠١٨	٢٠١٩	٢٠٢٠
عدد الشركات	٤	٤	٤	٤	٤	٥	٥	٥	٥	٦

عدد شركات التأمين التي تقدم تأمين البترول في السوق المصري

ملحق (٢)

نتائج أعمال فرع تأمين البترول بشركة مصر للتأمين عن الفترة من ٢٠١١ إلى ٢٠٢٠

القيمة بالألف جنيه

البيان السنة	صافي الأقساط	مخصص الأخطار السارية أول المدة	مخصص الأخطار السارية أخر المدة	الأقساط المكتسبة	صافي التعويضات	مخصص التعويضات تحت التسوية أول المدة	مخصص التعويضات تحت التسوية أخر المدة	التعويضات التحميلية	فائض / عجز النشاط التأمينى
٢٠١١	١٩٣٧٩٠	٨٧١٧٨	٨٢٤٦٥	١٩٨٦٨٢	٣٠٧٥٧	٩٤٨٣١	١١٤٩٢٦	٥٠٨٥٢	١٢٩٠٢٩
٢٠١٢	٢٣٦٩٠٦	٨٣٦٦٤	٩٠٨٧٢	٢٢٩٦٩٨	٥٠٨٠١	١١٦٥٨٤	١٨٢٥٢٧	١١٦٧٤٤	٤٧٨٩٦
٢٠١٣	٣٣٦٠٠٦	١٠٥٥٠٤	١٤٦١١٨	٢٩٥٣٩٣	٤٧٣٩٨	٢٠٥٤٨٧	٢٩٠١٠٠	١٣٢٠١١	٩٢١٩٦
٢٠١٤	٣٥٢٨٧٥	١٤٨٩٦٨	١٦٠٧٣٠	٣٤٠٨٤٣	١٤١٨٠٥	٢٩٤٤٨٩	٤١١٣٦١	٢٥٨٦٧٦	٣٦٦٢١
٢٠١٥	٣٥٢٨٧٥	١٦٨٤٩١	١٦٧٤٢٤	٣٣٨٣٥٤	١٤١٨٠٥	٤٢٩٣٢٢	٣٧٥٧٦٦	٩٤١٦٥	١٧٨٢٦٧
٢٠١٦	٣٦٦٨٤١	١٩١٥٧٢	١٥١٧٥٦	٤٠٦٦٥	٢٢٦٩٣٨	٤٢٢٢٩٦	٤٢٦٢٩٩	٢٣٠٩٤٢	١٣٣٨٢١
٢٠١٧	٥٣٧٣٥٦	٣٠٥١٣٩	٣٥٨٢٩٥	٤٨٤٢٠٠	١٦٨١٣٤	٥٩٦٤٧١	٦٣٧٦٨٥	٢٠٩٣٤٨	٢٧٠٥٢٩
٢٠١٨	٦٧٢٠٩٨	٣٥٤٢٨٨	٤٠٤٠٤٧	٦٢٢٣٣٩	٢٥١٩٢١	٦٣٣٠٥٩	٧٤١٧٦٨	٣٦٠٦٣٤	٢٥٧٢٤١
٢٠١٩	٧٧٤٣٢٩	٣٧٨١٣٣	٣٩٩٢٣٢	٧٥٣٢٣١	٣٢٧٢٧١	٧١٥٢٧٠	١٠٣٩٧٤٦	٦٥١٧٦٥	١٩٠٣٩٧
٢٠٢٠	٣٩٠٥٥٩	٣٨٥٨٧٧	١٩٢٥١٢	٥٨٣٩٢٤	٢٤٦١٣٦	١٠١٨٦٩٦	١٢٨٠٠٣٢	٥٠٧٤٧٢	٢٤٨٥٤٩

ملحق (٣)

نتائج أعمال فرع تأمين البترول بشركة قناة السويس للتأمين عن الفترة من ٢٠١١ إلى ٢٠٢٠

القيمة بالألف جنيه

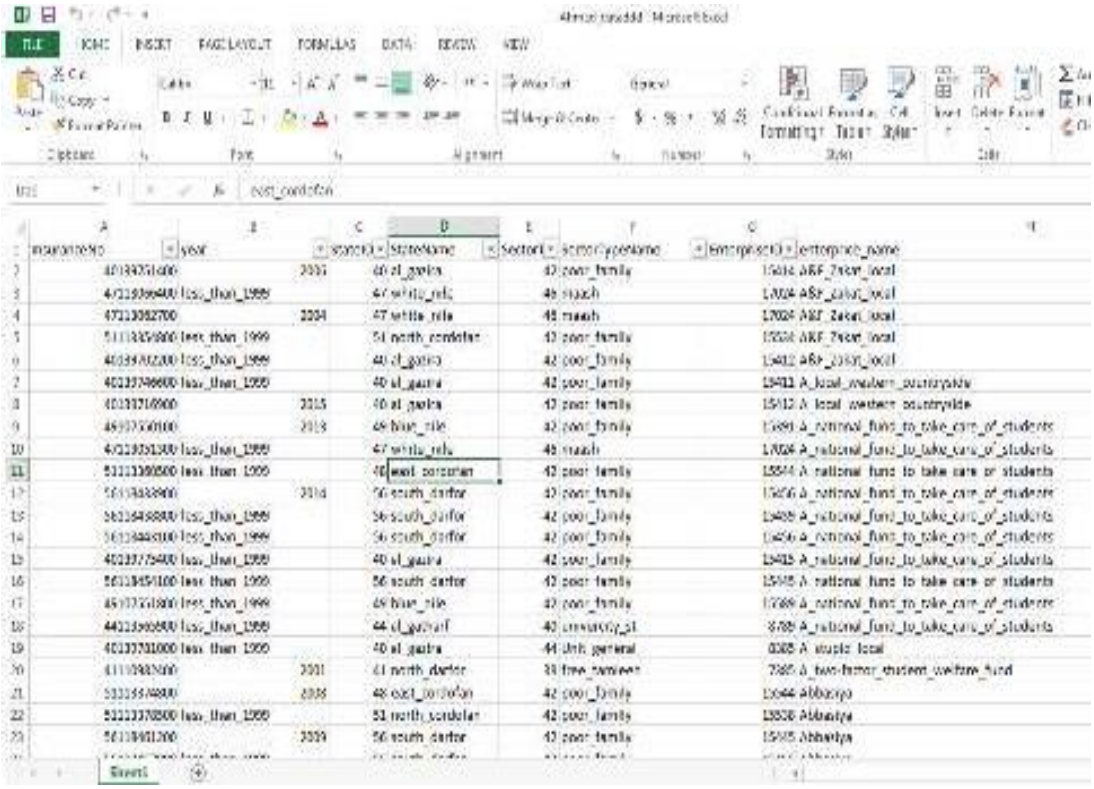
السنة	صافي الأقساط	مخصص الأخطار السارية أول المدة	مخصص الأخطار السارية أخر المدة	الأقساط المكتسبة	صافي التعويضات	مخصص التعويضات تحت التسوية أول المدة	مخصص التعويضات تحت التسوية أخر المدة	التعويضات التحميلية	فائض / عجز النشاط التأميني
٢٠١١	–	–	–	–	–	–	–	–	–
٢٠١٢	–	–	–	–	–	–	–	–	–
٢٠١٣	–	–	–	–	–	–	–	–	–
٢٠١٤	–	–	–	–	–	–	–	–	–
٢٠١٥	–	–	–	–	–	–	–	–	–
٢٠١٦	–	–	–	–	–	–	–	–	–
٢٠١٧	٩٤	–	٩٤	–	–	–	–	–	(١٤٩)
٢٠١٨	٣٠٩	٩٤	٣٠٩	٩٤	–	–	–	–	١٠٥
٢٠١٩	٥٢٧	٣٠١	٥٢٧	٣٠١	٢٥٦	–	–	٢٥٦	١٠
٢٠٢٠	٤٥٣	٥١٩	٢١٨	٧٥٤	–	–	٥	٥	٦٦٤

ملحق (٤)

واجهة برنامج WEKA

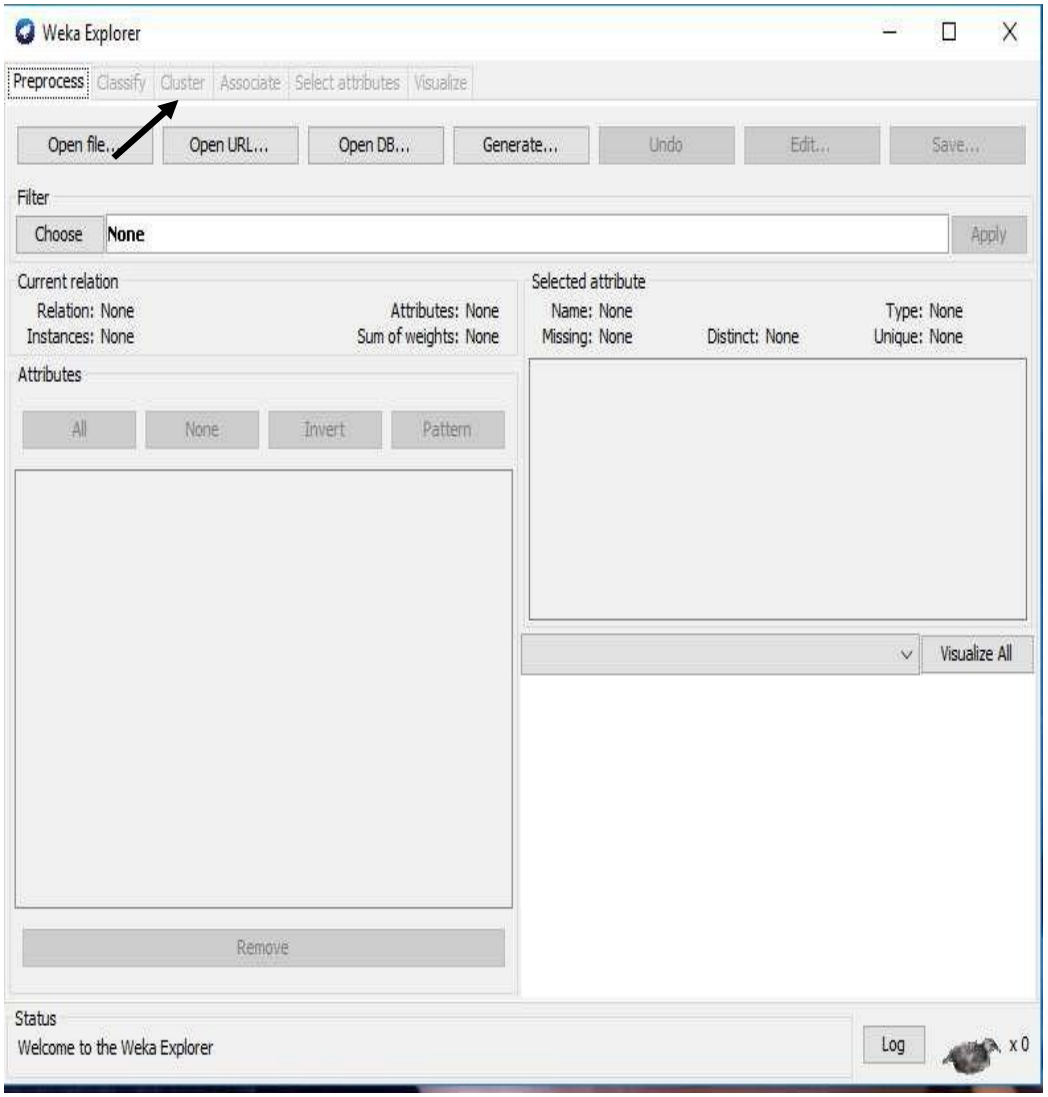
ملحق (٥)

جدول إدخال البيانات من خلال برنامج Excel



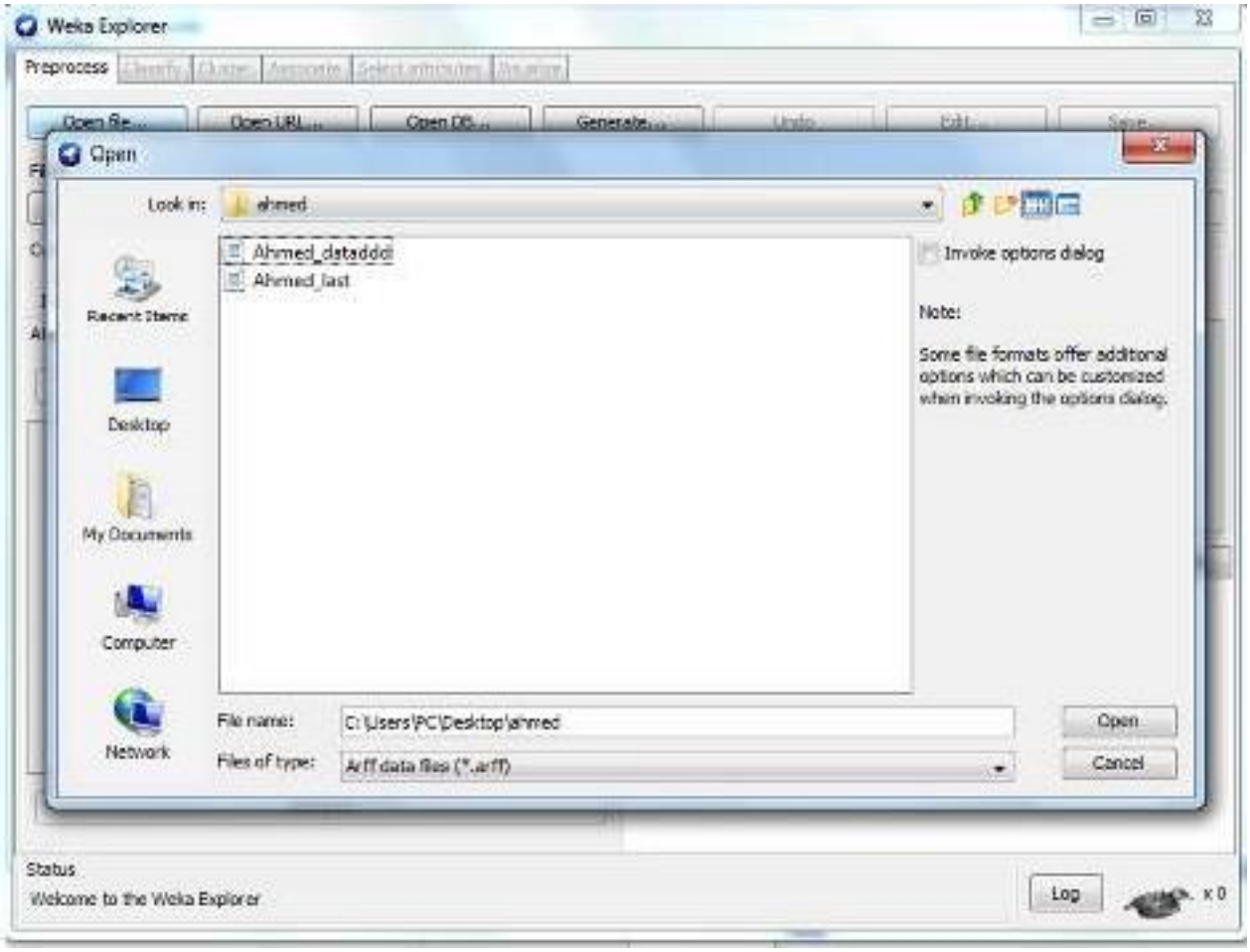
ملحق رقم (٦)

واجهة تحميل البيانات على برنامج WEKA



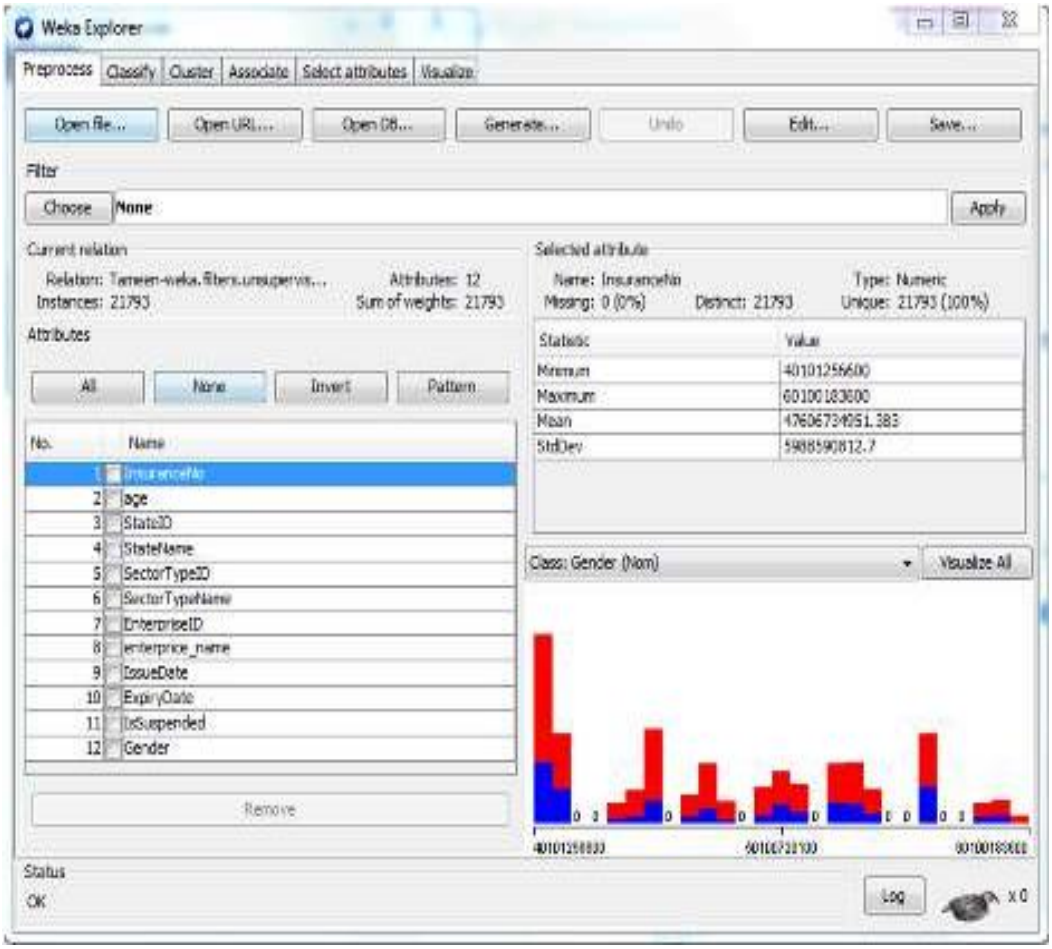
ملحق رقم (٧)

واجهة فتح الملف في برنامج WEKA



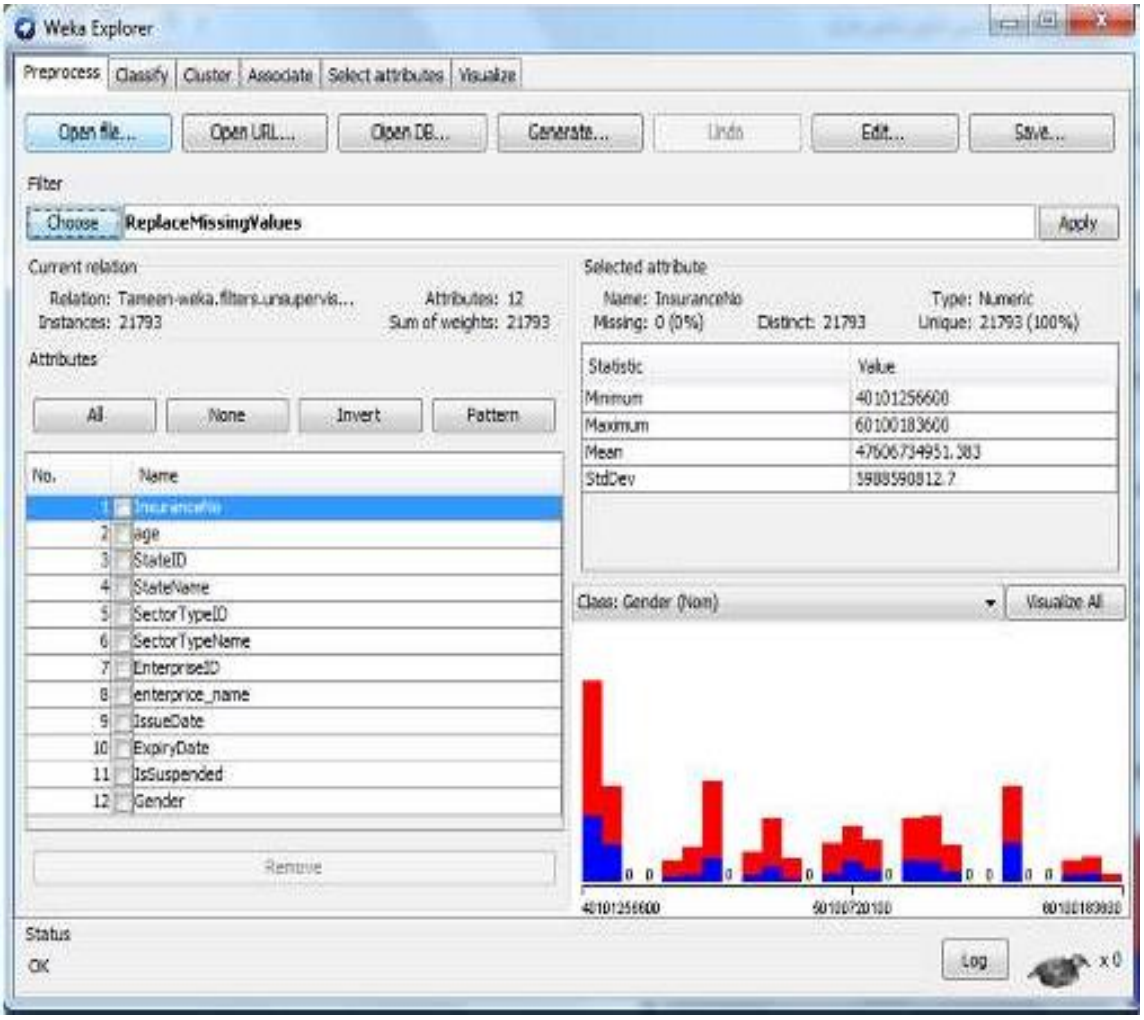
ملحق (٨)

واجهة تدريب البيانات على برنامج WEKA



ملحق (٩)

واجهة البيانات بعد تدريبها وتنظيفها على برنامج WEKA



ملحق (١٠)
مصفوفة الأداء بالنسبة لشركة مصر للتأمين

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	<-- classified as
898	27	9	21	8	42	20	68	52	3	13	15	7	29	46	151	24	1	a = al_gazira
90	123	4	7	2	28	5	33	29	5	5	1	0	8	17	97	5	0	b = white_nile
60	4	35	2	6	15	4	15	4	1	2	8	1	14	9	86	2	2	c = north_cordofan
65	2	5	72	11	17	2	8	7	1	4	3	0	5	10	52	5	5	d = blue_nile
40	5	3	9	39	4	1	14	1	3	2	3	0	3	5	53	3	0	e = east_cordofan
97	11	8	5	6	340	8	34	10	5	3	10	1	13	21	88	19	0	f = south_darfor
96	4	3	10	3	11	35	10	5	0	0	3	2	14	12	43	3	0	g = al_gatharf
174	19	8	5	3	26	7	301	16	8	5	7	1	20	33	91	15	0	h = north_darfor
21	4	0	0	0	7	1	4	99	0	1	0	0	5	7	16	4	0	i = west_darfor
56	6	0	3	2	13	0	17	2	12	3	6	1	6	5	35	4	0	j = sinar
54	7	3	9	4	9	4	19	9	2	31	2	1	9	9	57	7	0	k = Nile_rever
44	10	0	8	2	15	6	11	2	1	1	49	0	11	8	69	4	5	l = south_cordofan
45	6	4	3	2	8	5	19	4	2	4	1	17	12	17	61	2	0	m = Darfor_center
90	10	7	8	2	9	3	24	6	0	2	4	0	152	28	100	12	0	n = al_shmalia
52	9	7	2	1	9	6	40	3	5	1	6	1	28	136	79	13	0	o = east_darfor
92	11	5	20	7	19	3	57	10	0	5	7	1	24	10	427	5	4	p = Red_sea
111	4	1	6	0	25	3	26	17	1	3	5	2	32	22	71	119	0	q = Kasala
9	0	0	1	0	5	2	0	0	2	0	1	0	1	0	8	1	39	r = Abee_managment

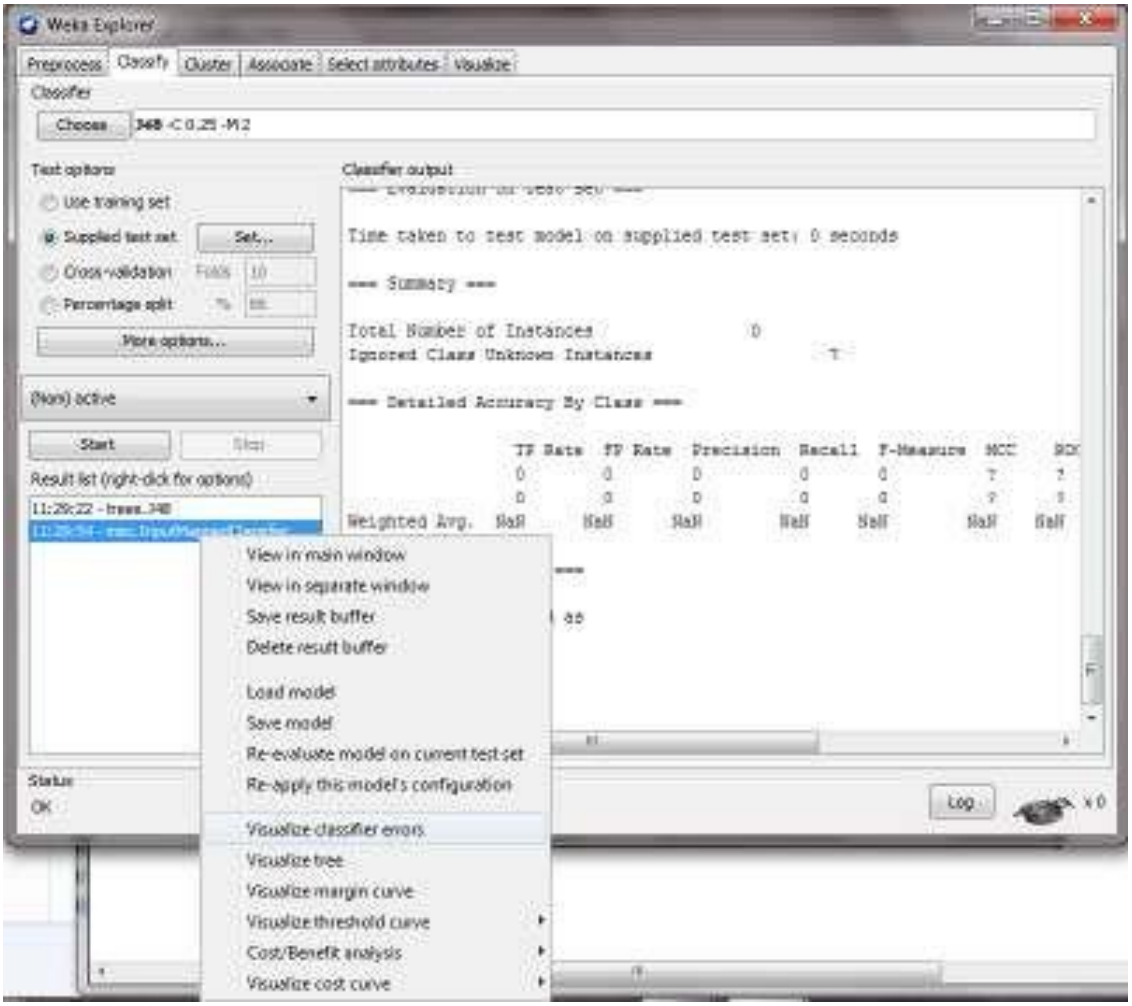
ملحق (١١)
مصفوفة الأداء بالنسبة لشركة قناة السويس للتأمين

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
3008	4	48	46	24	176	0	2	0	54	2	0	73	0	0	0	0	a = poor_family
87	1	11	0	4	17	0	0	0	16	0	0	1	0	0	0	0	b = maaah
322	0	249	0	12	112	0	2	0	28	0	0	20	0	0	0	0	c = univercity_st
99	0	15	93	3	32	0	1	0	15	1	0	1	0	0	0	0	d = Unit_general
420	1	21	11	44	75	0	0	0	29	0	0	15	0	0	0	0	e = free_tamleen
691	4	61	22	13	458	0	0	0	38	3	0	25	0	0	0	0	f = state_general
64	0	4	0	0	8	0	0	0	4	0	0	2	0	0	0	0	g = private
132	0	14	0	3	33	0	3	0	11	0	0	0	0	2	0	0	h = maaah_s_g
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	i = silf_servers
62	0	4	0	1	5	0	3	0	122	0	0	0	0	0	0	0	j = farmer
61	0	11	1	1	5	0	0	0	11	1	0	0	0	0	0	0	k = Itam_And_Aramil
5	0	0	0	0	2	0	0	0	5	0	0	6	0	0	0	0	l = shohada_family
19	0	0	0	2	6	0	0	0	0	0	0	223	0	0	0	0	m = school_st
20	0	0	0	0	5	0	0	0	1	0	0	0	0	0	0	0	n = Quran_st
1	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	o = hrfoan
3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	p = lowers
1	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	q = childeren

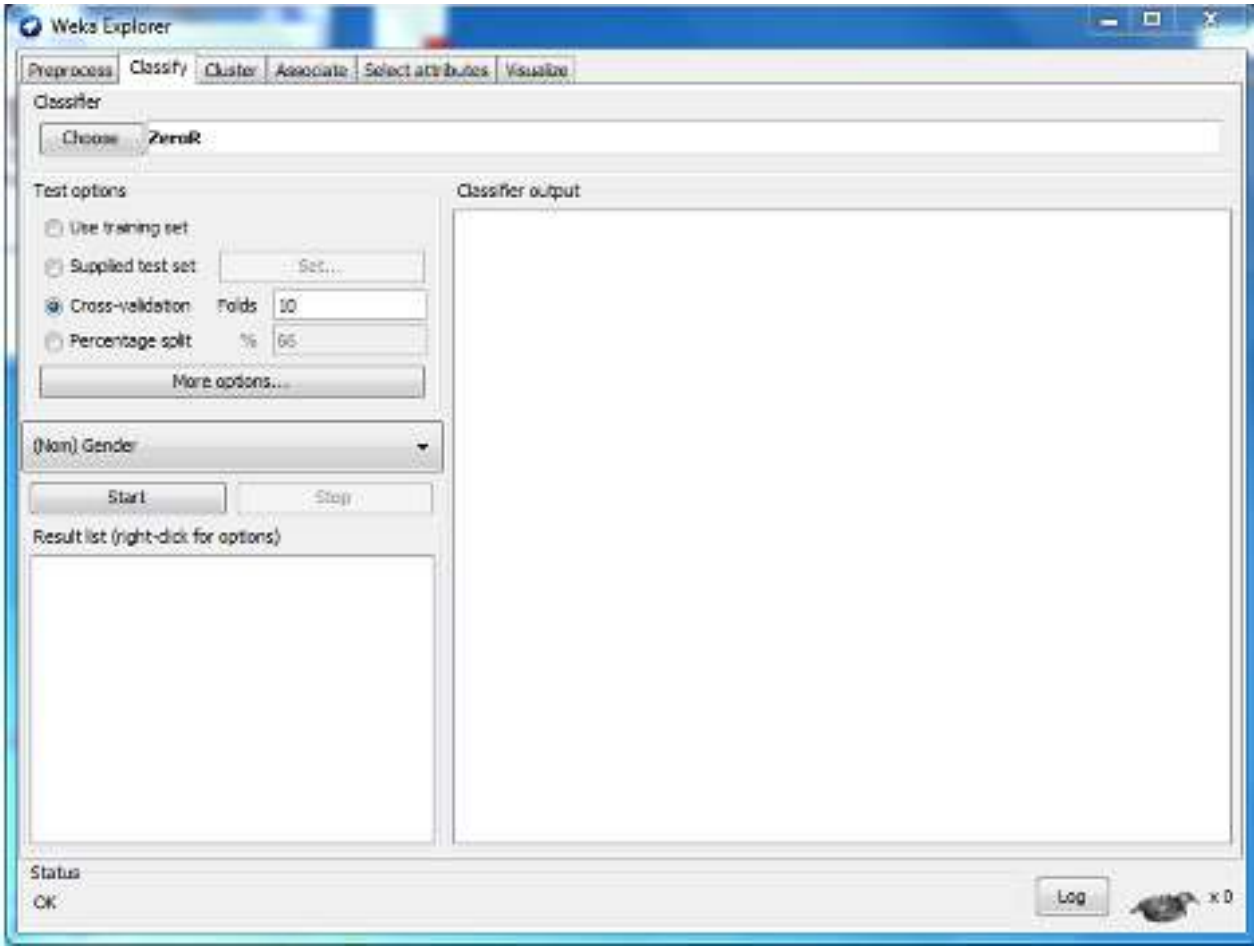
ملحق (١٢)

واجهة شاشة تطبيق الخوارزميات على برنامج WEKA



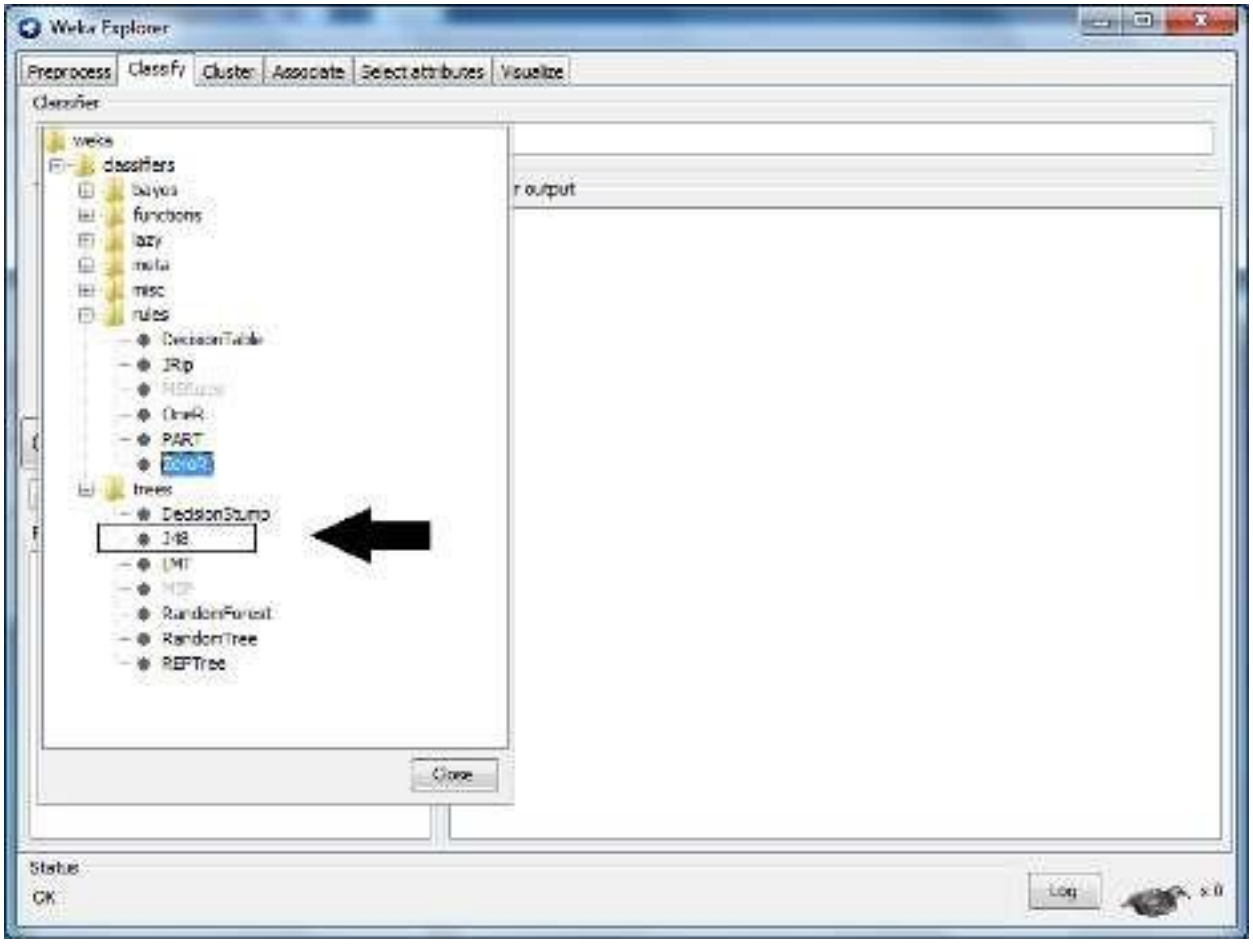
ملحق (١٣)

واجهة تنفيذ شجرة القرار في برنامج WEKA



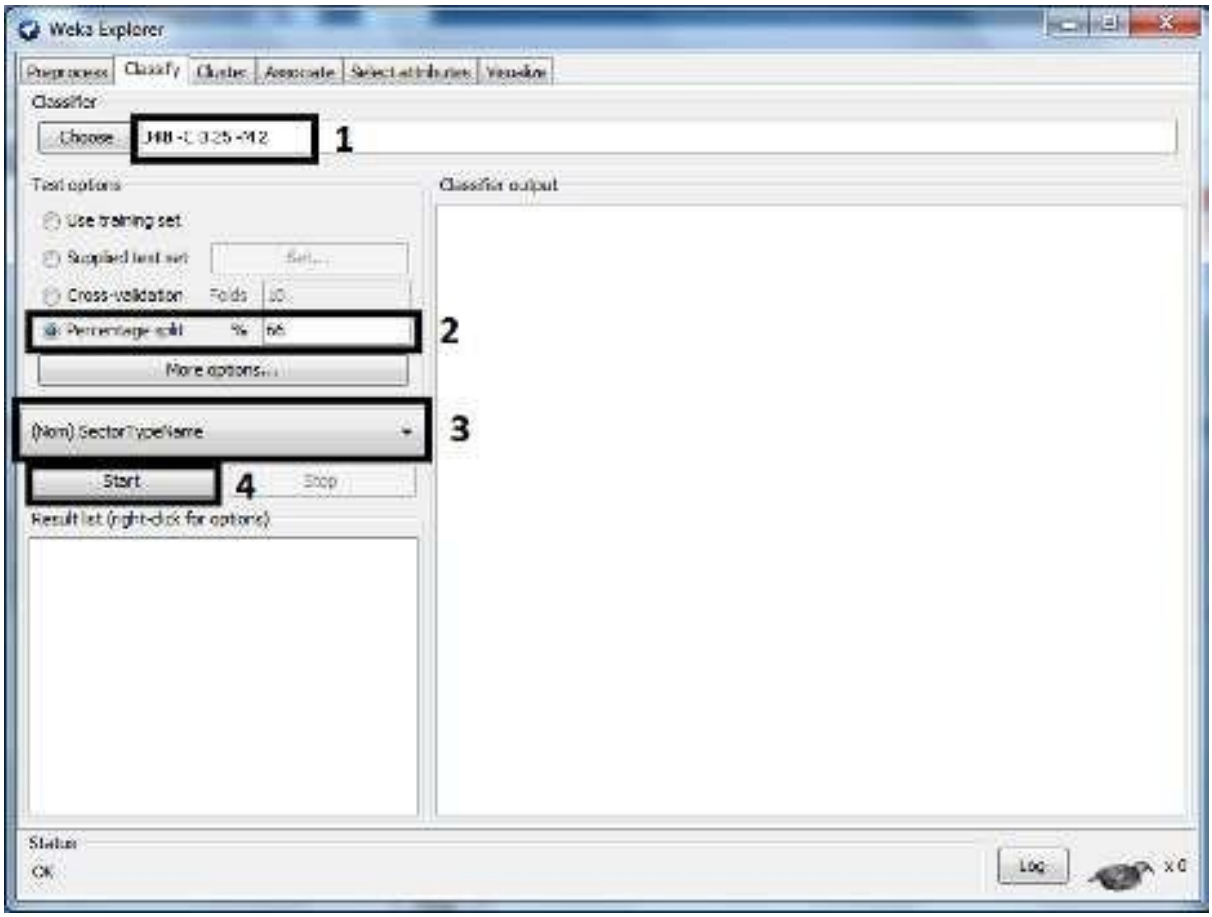
ملحق (١٤)

واجهة اختيار الخوارزمية في برنامج WEKA



ملحق (١٥)

واجهة ضبط الخوارزمية في برنامج WEKA



ملحق (١٦)

واجهة نتيجة تنفيذ الخوارزمية في برنامج WEKA