الفصل الاول

٢-١مـــشكـلـــة البـحـــث:-

هي وجود كمية كبيرة من البيانات دون الاستفادة منها وصعوبة في استرجاع البيانات وايضا ضياع الوقت في عملية البحث عن الملفات مما يعرض المؤسسة الى نقص في المعلومات الدقيقة التي يعتمد عليها .

٣-١اهمـــــيـــة البـــحـــث :-

تكمن اهمية البحث في ايجاد معلومات معينة في المعلومات الكبيرة مثلا عمليات التنقيب في المجالات الطبية والمجالات المصرفية والمجالات التربوية وغيرها من المجالات الاخرى.

١-٤ اهداف البحث :-

هناك ثلاثة أهداف للتنقيب في البيانات:

١) من أجل تعليل بعض الظواهر المرئية. مثال:

٢) من أجل التثبت من نظرية ما. مثال: التثبت من النظرية التي تقول بأن الأسر الكبيرة تهتم بالضمان الصحي أكثر من الأسر الصغيرة عددا.

٣) من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة. مثال: كيف سيكون الانفاق العام إن كان ملازما لعمليات خداع واسعة من قبل البطاقات الائتمانية.

ومن أحد أهداف تنقيب البيانات أيضاً هو اختزال كميات البيانات الكبيرة أو ضغطها بحيث تعبر بشكل بسيط عن كامل البيانات بدون تعميم.

١-٥ حدود البحث :-

١. حدود مكانية

جامعة ديالى كلية التربية الاساسية والشبكة العنكبوتية

حدود زمانية

٢٠١٨-٢٠١٩م

مصطلحات البحث:-

التنقيب عن البيانات، البيانات الكبيرة، وسائل التنقيب، تقنيات التنقيب.

١-١المقدمة ....

التنقيب عن البيانات هي عملية بحث محوسب ويدوي عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة. كما ويعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة) لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات وعادة يتم التنقيب في بيانات تكون جزءاً من كامل البيانات حيث يكون الغرض عادة تعميم النتائج على كامل البيانات (مثلاً تحليل البيانات الحالية لمستهلكي منتج ما بغرض توقع طلبات المستهلكين المستقبلية).

وللتنقيب نوعان هما

التنقيب الوصفي والتنقيب الاستشرافي

ويوجد ايضا للتنقيب طرق ونسق في البيانات ومنها :

١- طريقة الوصف

٢- طريقة التنبؤ

وايضا هناك وسائل لتنقيب البيانات

حيث هناك عدة وسائل مختلفة من أجل التنقيب في البيانات و اختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها. ويمكن اجراء عملية التنقيب في البيانات بالمقارنة مع سوق البيانات ومخزن البيانات.

بعض هذه الوسائل هي :

١- الاستدلال المبني على الحالات السابقة

٢- الكشف عن قانون

٣- معالجة الإشارة

٤-شبكات عصبونية

٥- منحنيات غير ثابتة

وايضا هناك عدة مراحل متتابعة لعملية تنقيب البيانات حسب الشكل التالي:

وتوجد عدة تطبيقات لتنقيب و تشمل بعضا من الأمثلة:

١ تحليل النزعة التجارية
٢- تحليل الاستعمال
٣-التسويق لفئة معينة
٤- كتابة تقرير مختصر عن فئة معينة
٥- فعالية الحملة
٦- جاذبية السلعة

الفصل الثاني

٢-١ تعريف التنقيب في البيانات:

هي عملية بحث محوسب ويدوي عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة. كما ويعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة) لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطلق اسم "نماذج" models على العلاقات والبيانات الملخصة التي يتم الحصول عليها من التنقيب في البيانات. يتعامل تنقيب البيانات عادة مع بيانات يكون قد تم الحصول عليها بغرض غير غرض التنقيب في البيانات (مثلاً قاعدة بيانات التعاملات في مصرف ما) مما يعني أن طريقة التنقيب في البيانات لا تؤثر مطلقاً على طريقة تجميع البيانات ذاتها. هذه هي أحد النواحي التي يختلف فيها التنقيب في البيانات عن الإحصاء، ولهذا يشار إلى عملية التنقيب في البيانات على أنها عملية إحصائية ثانوية. يشير التعريف أيضاً إلى أن كمية البيانات تكون عادة كبيرة، أما في حال كون كمية البيانات صغيرة فيفضل استخدام الطرق الإحصائية العادية في تحليلها.

عند التعامل مع حجم كبير من البيانات تظهر مسائل جديدة مثل كيفية تحديد النقاط المميزة في البيانات، وكيفية تحليل البيانات في فترة زمنية معقولة وكيفية قرار ما إذا كانت أي علاقة ظاهرية تعكس حقيقة في طبيعة البيانات. عادة يتم التنقيب في بيانات تكون جزءاً من كامل البيانات حيث يكون الغرض عادة تعميم النتائج على كامل البيانات (مثلاً تحليل البيانات الحالية لمستهلكي بغرض توقع طلبات المستهلكين المستقبلية). وتعبر عملية تنقيب البيانات بشكل بسيط عن كامل البيانات بدون تعميم عملية انتزاع الانماط المخفية من البيانات .

كي نفهم ما هو الذي يحصل في التنقيب سوف نوضح الفرق بين عملية الاستعلام العادية على قواعد البيانات العادية والاستعلام في البيانات المنقبة ..

المثال التالي والذي يوضح الفرق في الاستعلام بين كلا النوعين ..

الاستعلام عن معلومات في سوق تجاري ....
في قاعدة البيانات العادية (Data base)

* الاستعلام عن المعلومات الكاملة للزبائن الذين ينتهي اسمهم الثاني بخالد ..
* الاستعلام عن الزبائن الذين اشتروا بمبلغ اكثر من ١٠٠٠ دينار في اخر الشهر.
* الاستعلام عن اسماء الزبائن الذين اشتروا الحليب .
* الاستعلام عن كل البطاقات الائتمانية والتي عليها مشاكل .. وهذا يسمى التصنيف .
* الاستعلام عن الزبائن الذين لديهم عادات شرائية مماثله .. وهذا يسمى التجمع او العناقيد .

* الاستعلام عن السلع التي يتم شراءها بشكل تزامني مع الحليب ، المقصود في كل عملية شراء يقوم بها الزبائن .. وهذه تسمى اكتشاف قواعد وعلاقات الارتباط .

٢-٢وسائل التنقيب عن البيانات:

اختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها. يمكن اجراء عملية التنقيب في البيانات بالمقارنة مع سوق البيانات ومخزن البيانات.

بعض من هذه الوسائل هي: (يتبعها وظيفة كل وسيلة)

١- الاستدلال المبني على حالات سابقة : التفكير واستخلاص النتائج والقوانين من أمثلة حية و قضايا تم حلها سابقا.

٢- الكشف عن قانون : البحث عن منوال معين أو علاقة معينة في جزئية كبيرة من البيانات.

٣- معالجة الإشارات : ايجاد الظواهر المتشابه مع بعضها البعض

٤- شبكات عصبونية : تطوير نماذج قابلة لتنبؤ النتائج. هذه النماذج تم تطويرها بناء على أسس تم استنباطها من عقل الإنسان.

٥- منحنيات غير ثابته: ضياع المعلومات.

٢-٣ انواع التنقيب

هناك نوعان أساسيان للتنقيب في البيانات هما:

١- التنقيب الاستشرافي

التنقيب الاستشرافي ينتج عنه نموذج عن النظام الذي تصفه البيانات المستخدمة في التنقيب.

٢- التنقيب الوصفي

وهو التنقيب الذي ينتج عنه معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب

٢-٤ المراحل الرئيسية لعملية تنقيب البيانات

تتكون عملية تنقيب البيانات من عدة مراحل متتابعة حسب الشكل التالي:

١- المرحلة الأولى: جلب البيانات

في هذه المرحلة يتم تحديد مصادر البيانات وكيفية دمجها ببعضها في حالة كانت من أكثر من مصدر.

٣- المرحلة الثانية: تنظيف البيانات

٤- تعتبر هذه المرحلة من المراحل المهمة جداً والتي تتضمن تهيئة البيانات لعملية التنقيب. بعض البيانات تحتوي على قيم فارغة، وبعضها يحتوي على أخطاء، ومنها ما هو غير منطقي. في هذه المرحلة يتم التعامل مع هذه المشاكل وحلها.

٣- المرحلة الثالثة: اختيار البيانات

في هذه المرحلة يتم اختيار البيانات المهمة في عملية التنقيب، ويتم استبعاد البيانات الغير مهمة، المُكررة والتي ليس لها علاقة. البيانات المكررة أو الغير مهمة ستؤثر على عملية تنقيب البيانات بالتأكيد، لذا يجب الانتباه لها.

٤- المرحلة الرابعة: تحويل البيانات

في بعض الأحيان نحتاج إلى تحويل شكل البيانات أو قيمها إلى شكل اخر أو قيم أخرى بهدف جعلها مناسبة للخوارزمية المستخدمة في التنقيب، فبعض الخوارزميات مثلا تتعامل مع قيم رقمية فقط، وفي هذه الحالة سيتم إما استبعاد البيانات النصية في المرحلة الثالثة أو تحويلها لقيم رقمية في هذه المرحلة.

٥- المرحلة الخامسة: تنقيب البيانات

هنا تتم عملية استخدام أحد الخوارزميات الذكية (مثل شجرة القرار، الشبكات العصبية ..الخ) والتي ستقوم بعملية تنقيب البيانات واستخراج المعرفة منها واستخدام المعرفة الناتجة في عملية اتخاذ القرار.

٦- المرحلة السادسة : التقييم
في هذه المرحلة يتم استخدام طرق تقييم خاصة لعملية التنقيب، وهدفها قياس درجة دقة العملية التي تمت باستخدام عينات اختبارية من البيانات.

٧- المرحلة السابعة: المعرفة هنا يتم عرض استخدام النتيجة النهائية لتنقيب البيانات في عملية اتخاذ القرار والاستفادة من المعرفة الناتجة عن ذلك.
المراحل السابقة تشكل مجتمعةً عملية تنقيب البيانات، وهي ليست بالضرورة في اتجاه واحد فقط، نستطيع بعد إتمام أي مرحلة الرجوع لأي مرحلة سابقة وذلك في حال وجدنا أن هناك خلل ما أو عدم دقة في النتائج أو حتى لأغراض التجربة، وهذا الأمر يعطي عملية تنقيب البيانات نوعاً من الديناميكية التي تسهل

٢-٥ تنقيب البيانات والعلوم الأخرى
يُعتبر تنقيب البيانات ملتقى الجهود المبذولة من الباحثين في عدة مجالات من المعرفة، والذي من خلاله يتم تطوير وبناء تقنيات تتعامل مع البيانات وأشكالها المتعددة وأنواعها المختلفة بهدف مواجهة المشكلات في مجالات مختلفة كالهندسة، الأعمال، الصناعة، الطب والعلوم. يجمع تنقيب البيانات بين عدة علوم كالإحصاء وتعليم الآلة وقواعد البيانات وتقنيات الإظهار المرئي، ويتجلى هذا الجمع في مراحل وخطوات تنقيب البيانات بدءاً من تجهيز وتجميع البيانات وحتى النتيجة النهائية والتي تختلف حسب

أهداف وأغراض التنقيب.

بعض العمليات والطرق في تنقيب البيانات تأتي من علم الإحصاء، مثل نظرية naïve bays و maximum entropy وتقييم احتماليات التنبؤ، ويختلف تنقيب البيانات عن الإحصاء في أنه لا يتعامل فقط مع البيانات التي تحتوي أرقام فقط، فهناك بيانات بأشكال مختلفة، كما أن تنقيب البيانات يختلف في نوعية الطرق التي يتعامل فيها مع البيانات، حيث تميل هذه الطرق لمجال

تعليم الآلة أكثر من الإحصاء الذي يتعامل مع الفرضيات وكم البيانات. تنقيب البيانات يستخدم بعض تقنيات تعليم الآلة مثل الشبكات العصبية وشجرة القرار، ويختلف هدف تنقيب البيانات عن هدف تعليم الآلة، فتعليم الآلة هدفه إعطاء الحواسيب القدرة على تنفيذ مهام يقوم بها البشر عبر تعليمها، بمعنى آخر: استبدال الدور البشري، ولكن تنقيب البيانات هدفه مساعدة الدور البشري ودعمه وليس استبداله .يتمثل الدور الرئيسي لقواعد البيانات في حفظ البيانات والحصول عليها عند الحاجة، بينما دور تنقيب البيانات يتمثل في القدرة على قراءة هذه البيانات وتحليلها للمساعدة في اتخاذ القرار المناسب. تقنيات الإظهار المرئي يتم استخدامها كأداة في مرحلة تحضير البيانات أو مرحلة ما بعد تنقيب البيانات لإظهار النتائج.

٢-٦ مهام تنقيب البيانات

مهام تنقيب البيانات تُعَبِر عن طبيعة ونمط المعرفة التي نُنقب عنها ويُحدد ذلك طبيعة البيانات والخوارزميات المستخدمة في التنقيب، وتنقسم مهام تنقيب البيانات إلى نوعين:

١- مهام تنبؤيه : هدف هذه المهام هو القيام بعملية تنبؤ قيمة سمة من سمات البيانات استنادًا لقيم سمات أخرى.

السمة التي يتم تنبؤ قيمتها تسمى Target Class، ويوجد نوعان لهذه المهام، الأول يُسمى التصنيف (Classification) والثاني يُسمى بالتنبؤ (Prediction).

٢-مهام وصفية : هدف هذه المهام هو القيام بعملية وصف لخصائص البيانات إما عن طريق تجميع البيانات المتشابهة في خصائصها ضمن مجموعات ويسمى هذا النوع بتحليل العناقيد ، أو عن طريق إيجاد روابط وقواعد علاقات بين البيانات ويسمى هذا بتحليل الاقتران وينتج عنه قواعد تسمى (Association Rules). يوجد نوع ثالث من المهام الوصفية ومهمته إيجاد البيانات الشاذة أو المشوهة ويسمى بتحليل القيم الشاذة.

٢-٧تهيئةالبيانات

١- تنظيف البيانات :- هنا يتم التخلص من بعض البيانات التي تحتوي على عيوب مان تكون مطبعيه ، او بيانات قديمة لا تفيد في الوقت الحالي وغير ذلك .

٢-البيانات المفقودة :- عمليه تنقيب البيانات من اهم متطلباتها ان تكوت البيانات كامله لا تحتوي على قيم مفقودة ، وهنالك طرق طبعا لتصحيح هذه البيانات ، مثل الوسط الحسابي وغيرها من العمليات (Missing value) .

٣-اشتقاق البيانات :- في بعض الاحيان يستوجب اشتقاق بعض الأعمدة فتساعدنا في الحصول على معلومات مفيدة

٢-٨ تطبيقات تستخدم في تنقيب البيانات

وهذه بعض من التطبيقات : (يتبعها مثال لكل تطبيق)

١- كتابة تقرير مختصر عن فئة معينة : تطوير وإنشاء تقارير موجزة عن الزبائن المهمين وعن بطاقات الائتمان.

٢- تحليل النزعة التجارية :ايجاد الأسواق ذات قدرات النمو القوية أو الضعيفة.

٣- التسويق لفئة معينة : ايجاد الزبائن من أجل منح التخفيضات لهم لسبب معين.

٤- تحليل الاستعمال : ايجاد منوال معين لاستعمال الخدمات والسلع

٥- فعالية الحملة : مقارنة استراتيجيات الحملات مع بعضها البعض من أجل ايجاد أكثرها فعالية وتأثيرا.

٦- جاذبية السلعة: ايجاد السلع التي تباع مع بعضها البعض.

٢-٩ تطبيقات التنقيب في البيانات بدأت تنمو بصورة كبيرة للأسباب التالية:

١- كمية البيانات الموجودة في مخزن البيانات وسوق البيانات تنمو بصورة أسية .
ومن أجل ذلك، فإن المستخدم يحتاج إلى أدوات متطورة من مثل التنقيب في البيانات من اجل استخلاص الفائدة والمعرفة من هذه البيانات.

٢- الكثير من أدوات التنقيب عن البيانات بدأت تظهر مؤخرا، وكل أداة أفضل من الأخرى.

٣-المنافسة الشديدة الموجودة في السوق تدفع الشركات إلى الاستفادة القصوى من البيانات التي بيدها. عمليات التنقيب في البيانات تفعل ذلك تماما.

الفصل الثالث

٣-١ التنقيب في البيانات وأدوات الاستعلام

ثمة فروق واضحة بين تحري البيانات وطرائق الاستعلام ولغات الاستعلام الشائعة التي توفرها معظم نظم إدارة قواعد البيانات، مثل لغة SQL. ففي حين صُممت لغات الاستعلام بهدف تقديم وسائل تمكن مستخدمها من تحديد البيانات التي يود استخراجها من قاعدة ما، واختبار الفرضيات التي وضعها بناءً على البيانات أو القيم المُستخرجة، صممت أدوات التنقيب في البيانات لتضع الفرضيات وتختبرها وتستنتج منها معلومات جديدة. ولتوضيح هذه الفكرة يمكن الموازنة بين هذين المنحيين في المثال الآتي: تريد إحدى شركات تصنيع الحواسيب معرفة معدلات بيع منتجاتها والوسائل الكفيلة بتحسين مردودها. فتبدأ بدراسة كل منتج من منتجاتها لتحديد معدل البيع في بعض مناطق العالم وتوازن بين هذه المعدلات لمعرفة مدى قبول المنتج في تلك المناطق. ولتحقيق ذلك يمكن استخدام إحدى لغات الاستعلام مثل SQL، ثم تدرس العلاقة بين متوسط دخل ا لفرد في كل دولة وعدد المنتجات التي تشتريها تلك الدولة سنوياً وهكذا، وفي كل مرة هناك فرضية يجري إثباتها أو نفيها. أما تقنيات التنقيب في البيانات فإنها تعنى بالإجابة على السؤال الأعم من هذه الأسئلة الفرعية مثل «هل لاقى منتج ما نجاحاً كافياً، وماهي العوائق التي تحول دون الحصول على النجاح المطلوب، وما هي طرق تذليل تلك العقبات؟». لا تتطلب أداة التحري في هذه الحالة أي فرضيات، إنما تقوم بعملية الفرز والتصنيف والتحليل لاكتشاف العلاقات والمعلومات المخفية التي لا يمكن ملاحظتها دوماً.

وعموماً يمكن القول إنه في حال معرفة المطلوب تماماً، يمكن استخدام لغات الاستفسار مثل SQL، أما إذا كان ما يريده المستخدم غير واضح، ولديه فرضيات أو معايير، فلا بد من اللجوء إلى تقنيات تنقيب في البيانات التي صار لها أهمية كبيرة لغموض المطلوب في أكثر الحالات.

٣-٢ أمثلة عن أهمية تطبيق تقنيات التنقيب في البيانات

يمكن بيان أهمية تطبيق تقنيات تحري البيانات في مثال بسيط يعرض وضع شركة توزيع، تتلقى طلبات شراء من المواطنين، تُدخل هذه الطلبات في نظام معلوماتي قبل إرسال المواد والبضائع التي تتضمنها هذه الطلبات إلى الزبائن. بعد العمل مدة طويلة بهذه الطريقة، سوف يتوافر لدى هذه الشركة قاعدة معطيات مهمة تحتوي على المعلومات التي تتضمنها طلبات الزبائن في تلك المرحلة، وسوف تكون هذه القاعدة مصدراً مهماً للمعلومات، تستفيد منه الشركة في اتخاذ القرارات المناسبة لعملها، ووضع خططها المستقبلية وتحديد مجموعات الزبائن التي تتعامل معها، ومن ثمّ تصنيفها وفقاً للمناطق السكنية، أو الفئات العمرية، أو المنتجات المطلوبة، أغيرها من المعايير البسيطة أو المركبة. كما تستطيع الشركة دراسة وسائل التسويق والدعاية التي تستخدمها، ومردود هذه الوسائل. فقد تكتشف الشركة أن عدد زبائنها الذين تعرفوا على

منتجاتها من الرسائل البريدية التي أرسلتها إلى سكان منطقة سكنية معينة لم يتجاوز نسبة ٤٪، وأن الربح الذي حققته من بيع منتجاتها لهذه الشريحة من الزبائن لا يغطي ما أنفقته في إرسال هذه الرسائل، ومن ثمّ لابد من إيجاد وسائل ترويج أكثر ملاءمة مثل الإعلان عن تخفيضات في أسعار بعض المنتجات أو توفير طرق تسديد تناسب مستوى دخلهم. كما يمكن تطبيق تقنيات أكثر تعقيداً مثل الشبكات العصبونية لمحاكاة عملية إرسال الإعلانات بالبريد وتجاوب الزبائن مع هذه الرسائل. تتعلم الشركة من المعلومات المسجلة حول ما جرى في الماضي ما يرشدها بعد ذلك إلى اختيار العناوين التي يمكن أن تتجاوب مع هذه الطريقة في الإعلان.

كذلك يمكن الانتقال إلى مستوى أكثر تعقيداً من دراسة الزبائن والعلاقة التي يمكن أن تكون بين مجموعات الزبائن ومجموعات المنتجات التي تسوقها الشركة؛ فالزبائن الذين يشترون اليوم ملابس لأطفالهم الرضع، قد يشترون بعد عشر سنوات الألعاب الإلكترونية لهؤلاء الأطفال.

يتبين من هذا المثال أهمية معرفة هذه الأنواع من القواعد وتطبيقها لإيجاد فرص تجارية أفضل، ومعرفة مدى فعالية طرق التسويق التي تطبقها وتدقيقها، ومقارنة هذه الوسائل وتحديد ما يناسب شريحة معينة من الزبائن.

لا يقتصر التنقيب في البيانات على التطبيقات من النوع السابق، وإنما يشمل طيفاً واسعاً من النشاطات، فمثل هذا التقنيات تفيد، إذا ما طُبقت على قواعد البيانات التي تحوي معطيات عن المشافي و المرضى الذين يدخلون، والأدوية التي يتعاطونها وطرق العلاج المستخدمة، ومدى استجابة الأمراض لها، والتوزع السكاني للمرضى، وأحوال الطقس في هذه المناطق السكنية، والأعمال التي يمارسها سكانها، ومستواهم التعليمي في معرفة الأمراض التي تصيب فئة معينة منهم، واكتشاف الأوبئة التي تصيب الناس في أوقات معينة واتخاذ الاحتياطات اللازمة لها، وغير ذلك من المعلومات؛ تفيد في اتخاذ قرارات تتعلق بتحسين مستوى الخدمات الطبية التي تقدمها هذه المشافي.

ولقد ظهر مع التنقيب عن البيانات أدوات وبرمجيات تساعد في التنقيب عن الكم الهائل والمتزايد من البيانات للوصول إلى المعرفة في قواعد البيانات المختلفة، وتيسر هذه الأدوات العمل على معظم التخصصات العلمية ومنها علوم المكتبات والمعلومات. لذا تهدف هذه الدراسة إلى دراسة ماهية التنقيب عن البيانات ومهامه وتطبيقاته، وتحليل أدوات التنقيب عن البيانات مفتوحة المصدر وتقييمها، ومن ثم عقد مقارنة بين أدوات التنقيب عن البيانات مفتوحة المصدر. وتوصلت الدراسة للعديد من النتائج أهمها: هناك مزية يتصف بها بعض الأدوات والتي تتضح من خلال الاستخدام وهي توفير نموذج السحب والإفلات أثناء عملية التركيب والبناء للتنقيب عن البيانات وهي تتوافر بأربع أدوات . Rapid Miner، weka، Orange،KNIM.
rapidminer-logo-retina.png

Rapidminer –logo-retin.png هو منصة برمجيات طورت لتوفير بيئة متكاملة للتعلم آلة، واستخراج البيانات والتعدين النص، التحليلات التنبؤيه وتحليل الأعمال.

Orange مطورة بلغة البايثون وتتمير بسهولة الاستخدام (drag and drop) ويمكنك القيام بالعديد من العمليات مثل:

Data: data filtering, sampling, imputation, feature manipulation and feature selection

Visualize: box plot, histograms, scatter plot

Classify: a set of supervised machine learning algorithms for classification

Weka:هي أداة توفر العدد الكبير من خوارزميات تعلم الآلة وتعدين وتحليل البيانات مثل:

Clustering, Classification, Regression, Visualization, and Feature selection

الاستنتاجات

استنتجت الدراسة على ان الحصول على معلومة ما من معلومات او بيانات كبيرة لا يمكن عن طريق طرق تقليدية (البحث التقليدي) حيث يجب استخدام خوارزميات ذكية تتيح لنا الحصول على هذه المعلومات من

هذه البيانات الكبيرة لوقت قصير.

التوصيات

نوصي بتوسيع الدراسات العلمية على موضوع التنقيب لأهميته الكبيرة حيث العالم يشهد هذا التطور الهائل في التكنلوجيا والبيانات الكبيرة فالحصول على معلومة معينة من هذه البيانات يتم عن طريق التنقيب عنها باستخدام طرق عملية ذكية تتيح لنا الحصول على المعلومات لهذه البيانات.

الـــمـــــصـــــــادر.....

1-OKAIRP 2005 Fall Conference, Arizona State UniversityAbout.com: Data mining

2-rseichl, Markus (September–October 2011). ؛Markus, Ralf ^ "Data Mining 3-Tools". Wiley Interdisciplinary Reviews: Data Mining .and Knowledge Discovery. 1 (5): 431–445. doi:10.1002/widm.24 .October 21, 2011 اطلع عليه بتاريخ

4-Jerome (2009). ،Friedman ؛Robert ،Tibshirani ؛Trevor ،Hastie ^ "The Elements of Statistical Learning: Data Mining, Inference, and ."Prediction

٥ - ويكيبيديا wikipedia.org

٦- منتدى الفريق العربي arabteam2000-forum.com

٧- مدونة مبادئ إدارة نظم المعلومات nal111.wordpress.com

٨- منصة المعرفة www.marefa.org

٩- مدونة نسيج blog.naseej.com

https://www.nmthgiat.com/%D8%AA%D9%86%D9%82%D9%8A%D8%A8-%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7/%D8%AA

١٠- Data Mining Concepts and Techniques Second Edition.

١١- مقالة لـ (إيمان بن سالم و محمد خير الدين خلادي) – جامعة منوري قسنطينة.

مرسلة بواسطة (teamwork) في ٣:٢٦ ص .