جامعه كردفان

كليه دراسات الحاسوب والاحصاء

مقترح بحث تكميلي لنيل درجه بكلاريوس الشرف في علوم الحاسوب

بعنوان:

استخدام تقنيات التنقيب في البيانات باستخدام الweka

اعداد الطلاب:

١- محمد سعيد العبيد بشاره
٢- عبد السميع اسماعيل محمد اسماعيل
٣- أنس محمد عبد المتعال عبد الرحيم
٤- عمر مصطفي النعيم محمد
٥- التوم عبد الباقي التوم
٦- احمد الرحيمه احمد الطاهر
٧- مجاهد يوسف عمر عثمان
٨- عرفه مردس سليمان محمد
٩- مواهب الامين ابوبكر الصديق
١٠- صفاء الطيب سعيد حبيب

اشراف الاستاذ:

مجتبي فضلي المولي الصافي

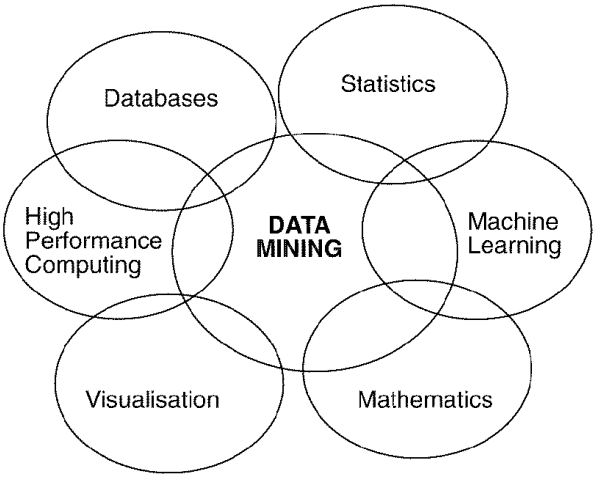
مقدمه:

من المؤكد أن التنقيب في البيانات Data mining هو من الموضوعات بالغة الأهمية ... وإلا ما الذي يجعلها أحد الموضوعات الأكثر أهمية لرئيس الولايات المتحدة أوباما ؟ فالمتتبعون لنشاط رئيس الدولة الأكثر تحكماً في العالم سيعرفون أنه بتاريخ ١٧ ديسمبر ٢٠١٣، كان في لقاء مع قيادات تقنية المعلومات والذين كان منهم Tim Cook من شركة أبل Apple و Eric Schmidt من جوجل وكذلك المدراء التنفيذيين لكل من Twitter، Microsoft، Face book، Sales force، Netflix، Drop box، Yahoo، Sherpa Global، Comcast، LinkedIn وذلك لمناقشة موضوع التنقيب في البيانات . كما أنه في خطاب له بتاريخ ١٧ يناير ٢٠١٤ عاود من جديد طرح الموضوع من خلال الدعوة إلى إجراء إصلاحات في نظام التنقيب عن البيانات.

ذلك يؤكد الاهتمام الذي يتم إيلاءه لموضوع التنقيب في البيانات على المستوي الدولي والعالمي.

أدي الانتشار الواسع لتقنية المعلومات وسهولة إتاحتها إلى تضخم حجم المعلومات بصورة استباقية لم يشهدها التاريخ من قبل، مما جعل من قضية البيانات الضخمة على الإنترنت مثاراً للجدل، من حيث جدوى وجودها بهذه الصورة العشوائية. وعندما نتحدث عن البيانات الضخمة، فإننا نتحدث عن كميات لا يمكن تخيلها من البيانات متعددة الأنواع والمصادر بحجم يصل إلى المئات من التيرابايت أو حتى البيتابايت (البيتابايت هو الرقم واحد متبوعا بـ ١٥ صفر). كما أن شركة IBM تتحدث عن ٢.٥ كوينتيليون بايت من البيانات كل يوم (الكوينتيليون هو الرقم واحد متبوعا بـ ١٨ صفرا). معلومات في كل مكان وفي موضوعات شتي. وهنا نتساءل ما أهمية هذه البيانات في ظل وجود حقيقة تشير إلى أن المعلومات المنظمة من هذه البيانات لا تشكل إلا جزءً ضئيلاً يصل إلى ١٠% مقارنه بالمعلومات غير المنظمة والتي تشكل الباقي. ذلك أدي إلى ازدياد الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، فالأساليب التقليدية والإحصائية لا تستطيع أن تتعامل مع هذا الكم من الهائل لذا تستخدم أدوات ذكية لمعالجة هذه البيانات.

من هنا ظهر ما يسمى باستخراج البيانات Data Mining كتقنية تهدف إلى استنتاج المعرفة من كميات هائلة من البيانات، تعتمد على الخوارزميات الرياضية والتي تعتبر أساس التنقيب عن البيانات وهي مستمدة من العديد من العلوم مثل علم الإحصاء والرياضيات والمنطق وعلم التعلم، والذكاء الاصطناعي والنظم الخبيرة، وعلم التعرف على الأنماط ،وعلم الآلة. وغيرها من العلوم والتي تعتبر من العلوم الذكية وغير التقليدية.



ظهر التنقيب في البيانات (mining Data) في أواخر الثمانيات وأثبت وجوده كأحد الحلول الناجحة لتحليل كميات ضخمة من البيانات، وذلك بتحويلها من مجرد معلومات متراكمة وغير مفهومة (بيانات) إلى معلومات قيِّمة يمكن استغلالها و الاستفادة منها بعد ذلك.

وقد اجتذبت مرحلة التنقيب في البيانات الكثير من الاهتمام في الأوساط البحثية على مدي العقد الماضي، في محاولة لتطوير خوارزميات قابلة للتوسع والتكيف مع كميات متزايدة من البيانات في البحث عن أنماط معرفية ذات معنى. وقد نمت حزم من الخوارزميات والبرمجيات و بشكل كبير خلال العقد الماضي، إلى حد أن التوسع قد جعل من الصعب على العاملين في هذا الحقل تتبع التقنيات المتاحة لحل مهمة معينة.

ونعني بالتنقيب في البيانات:

فيمكن ان نعرف (Data mining) هو عمليه بحت محوسب ويدوي عن معرفه من البيانات دون فرضيات مسبقه عما يمكن ان تكون هذه المعرفه.

تعريف اخر:

هو عباره عن تقنيه تستخدم في استخلاص المعرفه من قواعد البيانات

مشكله البحث:

١- عدم اختيار الخوارزميه الافضل من بين الخوارزميات التي تستخدم في تنقيب البيانات .

٢- عدم استخدام النسبه الصحيحه للخوارزميات التي تستخدم في تنقيب البيانات.

أهداف البحث:

١- معرفه الخوارزميه الافضل من بين الخوارزميات التي تستخدم في التنقيب في البيانات(Data mining).

٢- معرفه النسبه الملائمه التي يمكن استخدامها.

أهميه البحث:

تهدف اهميه البحث الي تقييم ومقارنه الخوارزميات المستخدمه في التنقيب في البيانات واختار افضل خوارزميه ونسبه من بين الخوارزميات المختاره ونسبها.

منهجيه البحث:

يستخدم البحث المنهج الوصفي والتحليلي في المصادر واستخدام برنامج(Weka)الذي يستخدم في التنقيب في البيانات.

هيكليه البحث:

*المقدمه

*مشكله البحث

*أهداف البحث

*أهميه البحث

*منهجيه البحث

هيكل البحث:

يحتوي البحث علي خمس فصول.