بسم الله الرحمن الرحيم

جامعة السودان للعلوم والتكنولوجيا
كلية الدراسات العليا

***

عنقدة بيانات السرطان بإستخدام خوارزمية الخطوتين

Cancer patients Data Clustering Using Two-Steps Algorithm

بحث مقدم كأحد متطلبات نيل درجة الماجستير فى علوم الحاسوب

مقدم البحث:
أحمد حمزة عثمان أحمد

المشرف:
د.محمد الحافظ مصطفى

سبتمبر 2008

الــحمد لله

الحمد لله الذى أعاننى على اتمام هذا البحث المتواضع ولما حصلت علية من معرفة وأحمد الله تبارك وتعالى أن تفضل علي بأن زودنى بما لم أكن اعرفة حتى تم إخراج هذا البحث الذي أهدف به بأن يُسهم في تطبيقات تقنيات التنقيب وذلك لايجاد نتائج يمكن ان تفيد المجتمع خاصة فى مجتمعنا السودانى لان هذا البحث أعتمد على البيئة السودانية فى مدخلاته فبالتالى مخرجاتة بلا شك كانت عائدة لنفس البيئة .

الأيـة

بسم الله الرحمن الرحيم

((قُل لَّو كَان البَحْرُ مِداداً لِكَلماتِ رَبِّي لَنَفِدَ البَحْرُ قَبْلَ أنْ تَنَفَدْ كَلِمَاتُ رَبَي ولَو جِئْنَا بمثْلِهَ مَدَدَاً))

صدق الله العظيم
سورة الكهف – الأية (109)

إهـــــــــــــداء

إلى من ربيانى وعلمانى دنيا الحياة والعلم والبحث عن الهدف
(والدتى ووالدى)

الى التى وقفت بجانبى ليل نهار وكان لها القدر المعلى فى إخراج هذا البحث(زوجتى)

الى كل العلماء الذين بذلوا جهودهم وارواحهم للتقدم والأزدهار

إلى كل من يبحث عن الجديد فى دنيا التكنلوجيا

إلى كل والد ووالدة يطمحان بأن يصل ابنائهم الى أعلى درجات العلم

إلى اساتذتنا الاجلاء واخواننا الباحثين والطلاب .....

اليكم ولكم جميعا أُهدى هذا البحث المتواضع سائل المولى ان تعم به الفائدة

شـــكر و عرفان

الشكر أجزله لله الواحد القهار. ..

ماطلعت شمس وما غاب نهار.....

الشكر أوفره لكل من نصب خيمة علمٍ كمنار.....

الشكر أوسعه لكل والدٍ ووالدة رفعوا شعار العلم شعار.....

الشكر أكمله لكل معلم ومعلمة ينهضون بالجبال لإزالة ماعلى عقولهم من غبار.....

الشكر خالصه لجامعة إفريقيا العالمية التى قدمت ومازالت تقدم العلم لطلابها أنهاراً انهار.. وأخص منهم الدكتور قسم السيد ابراهيم والأستاذ المربى الدكتور عماد تاج الدين ابراهيم

والشكر أجزله لأستاذى الدكتور (محمد الحافظ مصطفى) الذى ظل خلفى ليل نهار ولم يبخل علي بعلم وأفكار.....

والشكر لكل أساتذتى الذين تلقيت على أيدهم العلم فى جميع المراحل التعليمية والشكر كل الشكر للزملاء والزميلات بجامعة السودان الذين كنت معهم طيلة فترة الدراسة واخص بهم الصديق العزيز/ البراء ابو عبيدة ..رفيق دربى دراستاً وعملاً والزميلة ندى محمد عثمان والزميل رامى سعد الدين والزميل أبوبكر الصافى

كما أخص بالشكر ايضا الاخ الزميل الصديق/ عبد الماجد محمد والاخ عبد الرحمن أدريس والاستاذ النور عثمان الذين وقفوا معى فى إخراج هذا البحث
ولا انسى ان اشكر أسرة مستشفى الذرة وأخص منهم الاستاذه /ست النساء على ما أمدتنى به من معلومات للأكمال هذا البحث

المستخلص

علم التنقيب فى البيانات هو العلم الذى يبحث فى عملية الكشف عن معلومات ذات فائدة فى قواعد البيانات الكبيرة او أى بيانات كثيرة محفوظة بإى شكل .و من أهم عمليات التنقيب طريقة تحليل العنقدة وهى طريقة تهدف الى تقسيم البيانات الى تجمعات من البيانات وايجاد معلومات لم تكن معروفة من قبل او لم يهتم بها الخبراء فى الحقل.والعنقده هى تقسيم البيانات الى مجموعة من الاصناف اعتمادا على اشتراكها بالخواص المتشابهة . وتم تطبيق هذا البحث على بيانات مرضى السرطان حيث يعتبر المرض من الأمراض المنتشرة فى السودان , اما عن هذه البيانات التى تم التحصل عليها فكانت بيانات جيدة ولكنها تحتوى على قيم مفقودة وبعض القيم الغير منظقية فعمل البحث على تنظيف هذه البيانات حتى يتم الحصول على نتائج منطقية ومفيدة وجديدة فكانت النتائج منطقية من حيث التقسيم الى مجموعات فكانت هنالك عدد من التجارب فى عملية التقسيم , حيث تم التقسيم الى عشرة عناقيد وتسعة وثمانية الى ان حصلنا على مجموعتين ولكن هذه التقسيمات لم تظهر نتائج جديدة.وخرج البحث بعدد من التوصيات يمكن ان تظهر معلومات جديدة فيمكن إضافة هذه البيانات بزيادتها او استخدام خوارزمية عنقدة غير خوارزمية الخطوتين(Tow-Steps).

Abstract

Data mining is a science that relates for knowledge and gives useful information for large database or any amount of data saved in any manner.
One of important techniques of data mining called by Clustering method, this method produce some of groups or clusters, some of them include anomalies cluster or anomalies group, this cluster unknown by experts or workers in the hospital .
They are many data mining algorithms such as (Tow-Step) that is used to data cluster in ALTHRAA hospital in Khartoum.
The data collected in 2007 by International Health care group. This data talked Microsoft Excel format, and included to some problem such as missing values and outlier values, but this problem solved by Data Cleansing stage.
Data mining implemented about (2-10) clusters or groups but not absorbed any new discover Knowledge and information.
The data set implemented in own project included to some homogenous values like the Age or Sex attributes also included some heterogonous values such as Morphology or topography attributes .
All of this tested will be let to new discover Knowledge or information in near future, but I thing if increased the data set.
Data mining is a new science in computer field that is development the counters and companies to success.

فهرس الأشكال

فهرس الجداول

فهرس المحتويات