جامعة السودان للعلوم والتكنولوجيا
كلية علوم الحاسوب وتقانة المعلومات

# تقويم اداء الطالب الأكاديمي باستخدام تقنيات تنقيب البيانات

# Evaluation of student performance using data mining techniques

إعداد الطالبات :
١. امنية علي عجب.
٢. علياء عبد العزيز بشير.
٣. مشتهى ابراهيم محمد.

مشروع مقدم كأحد متطلبات الحصول على بكالوريوس الشرف في علوم الحاسوب

توقيع الأستاذ المشرف:
أ. وفاء فيصل مختار

اغسطس ٢٠١٤

# الأيـــــــة

(قل لو كان البحر مداداً لكلمات ربي لنفذ البحر قبل ان تنفذ كلمات ربي ولو جئنا بمثله مددا)

سورة الكهف الآية "١٠٩"

# الحمدلله

الحمدلله الذي لا يحمد على مكروه سواه، الحمدلله الذي به تتم النعم وتكتمل الاعمال، الحمدلله عدد ما كان وعدد ما يكون وعدد الحركات والسكون، الحمدلله في الاولى والاخرة.

ولا نرى في الحمد إلا زيادة في العطاء فنحمدك الله على برك الجزيل ونشكر فضلك على توفيقك لنا وإعانتنا في إتمام هذا البحث وتقديمة بالصورة المطلوبة.

ربي أدم صلاتك الكاملة وبركاتك الشاملة وسلامك الأتم على سيد الخلق اجمعين محمد خاتم النبيين "ﷺ" وعلى آله وصحبه والتابعين، وتابعيهم ومن اهتدى بهديه الى يوم الدين.

# الإهـــــــــــــــــــــــــداء

الى منبع الحنان ومعنى الأمان ونهر الحب الذي يجري بأرواحنا والقلب العطوف وجنتي في دنياي واخرتي،،،،،،امي الحبيبة

الى من احمل اسمه بكل افتخار وفية الهيبة والوقار ومن اعطاني بدون انتظار وعلمني العطاء والايثـــــار،،،،،، أبي الحبيب

الى من علمتني حروفا من ذهب وأسخت علي بالعلم والمعرفة ومنحتني نواقيس تضئ حياتي وطريقي،،،،،الاستاذة وفــاء فيصل

الى انقى القلوب واصفاها ،، من جمعنا دم واحد في العروق واسم واحد في الحياة،،،،،اخوتي واخواتي

الى رفقـــــــــــــاء المسيرة والدرب من فرحنا سويا وبكينا سويا وضحكنا سويا ورسمنا ملامح اربعة سنوات على دفاتر الزمن ولونها لنا القدر بفراق اجســـاد وبقاء ارواح وذكريـــــــــات،،،،،

اصدقاءنا الاوفياء .....

# الشـــــكر والتقـــــدير

تتناثر الكلمات حبراً وحباً.. على صفائح الاوراق.. لكل من علمني..

ومن أزال غيمة جهل مررت بها .. برياح العلم الطيبة .. ولكل من أعاد رسم ملامحي ..وتصحيح عثراتي..

أبعث تحية وشكر وإحترام إلى التي كانت الشمعة المضيئة لخروج هذا العمل على هذا النحو من التنظيم والترتيب في كل معلوماتة وكذلك شكله الخارجي الى الموجهه والمعلمة والدليل والنور في هذا البحث الاستاذة وفاء فيصل مختار..

والشكر أجزله موصول لجميع الاساتذة الأفاضل في كلية علوم الحاسوب وتقانة المعلومات والذين لم يبخلوا علينا بوقت أو معلومة ، ونخص بالشكر الاستاذة هدى جمال ، والاستاذة لبابة الطيب والشكر موصول الى الاستاذ عبد الحنان يس بالشؤون العلمية ، ولكل من ساعدنا من زملائنا وزميلاتنا بالكلية.

نسأل الله أن يوافيهم بأحسن ما عملوا ويجزيهم خير الجزاء.

# المستخلص

من المعروف أن سجل الطالب الأكاديمي هو الأساس الذي يرتكز عليه الطالب والمؤسسة التعليمية أثناء فترة الدراسة ، حيث أنه يحوي كل المعلومات الأكاديمية والشخصية حول الطالب علي مدار سنين الدراسة. وبالرغم من توفر كل المعلومات التي تساعد في إتخاذ القرارات بشأن تطوير المستوى الأكاديمي للطالب إلا أنه غير مستخدم لهذا الغرض و يكتفى منه بأخذ إحصائيات معينة وتخزينها.

وقد قمنا في هذا البحث بإستخدام إحدى تقنيات تنقيب البيانات وهي تقنية قوانين الأرتباط (Association Rules) وتطبيق إحدى خوارزمياتها وهي خوارزمية (Aproiri) في متابعة أداء الطالب إستناداً علي درجاته في كل المواد علي مدار الأربع سنين ودرجات الشهادة الثانوية وبعد تجميع البيانات وتهيأتها وتحليلها وتطبيق التقنية المذكورة مسبقا تم التوصل إلى علاقات توضح الإرتباط بين مواد معينة وتؤثر علي بعضها البعض .

# Abstract

It is known that the academic student's record represents the base upon any educational institution during the study period. It contains all the academic and personal information of the student over the years of his/her study, and despite the availability of all the information that will help in making decisions on the development of the academic level of the student, but it is not use for this purpose and it has just been used to take certain statistics and stored.

we have used in this search one of the popular data mining techniques which is called Association Rules, apply one of its most used algorithms which called Aproiri, in the follow-up of student performance based on grades in all subjects over the four years and degrees of secondary school certificate. We collected data of two academic years, prepare, analyze and apply of the aforementioned technology, we reached to relationships that describe the correlation between certain subjects and affect each other.

# المصطلحات الموحدة

| الإختصار | المصطلح | شرح المصطلح |
|---|---|---|
| DM | Data Mining | تنقيب البيانات |
| KDD | Knowledge Discovery in Data base | إكتشاف المعرفة المفيدة من مجموعة بيانات |
| | Confidence | قانون يقوم بقياس عدد مرات ظهور العنصر |
| | Support | وهذا مؤشر على مدى تكرار ظهور العناصر في قاعدة البيانات |
| | Item set | مجموعة العناصر التي تمثل معا كيان واحد |
| TID | Transaction identifier | يعرف المعاملات التي تتغير قيمتها |
| | Aproiri | خوارزمية للتنقيب عن البيانات |
| | Weka | برنامج للتنقيب عن البيانات(tool) |
| | Nominal | نوع بيانات (إسمي) |
| ROC | Receiver operating characteristics | رسوم بيانية مفيدة في تنظيم المصنفات وتصور الاداء |
| RARM | Rare Association Rule Mining | تقنية تقوم على اطار confidence-support وتمكننا من توليد قواعد متكررة على اساس frequent item set |
| ARM | Association Rule Mining | هي عبارة عن طريقة مدروسة وشائغة لأكتشاف العلاقة بين المتغيرات في قواعد البيانات الواسعة |
| ARFF | Attribute-Relation File Format | هي صيغة للعلاقات بين الصفات |
| | Numeric | نوع بيانات (رقمي) |
| | Explorer | زر في برنامج weka |
| CSV | Comma separated Value | صيغة لحفظ الملفات |
| | Large Item set | هو العنصر الأكثر تكراراً في المجموعة |

# فهرس الأشكال


| رقم الشكل | موضوع الشكل | رقم الصفحة |
|---|---|---|
| 1.2 | خطوات اكتشاف المعرفة في قواعد البيانات | ٥ |
| 2.2 | أنواع تنقيب البيانات | 5 |
| 1.4 | واجهة اللإختيار الرسومية في Weka | 21 |
| 2.4 | واجهة المستكشف في Weka | 22 |
| 3.4 | شكل شاشة Weka بعد ظهور النتائج | 23 |
| 1.5 | مخطط يوضح نسبة الذكور للإناث | 27 |
| 2.5 | الأداء في المواد التي تمثل مواد التخصص لقسم الحاسوب والشبكات | 27 |
| 3.5 | الأداء في المواد التي تمثل مواد التخصص لقسم الحاسوب والشبكات | 28 |
| 4.5 | الأداء في المواد التي تمثل متطلبات الكلية | 28 |
| 5.5 | الاداء في مواد متطلبات الكلية | ٣٢ |
| ٦.٥ | العلاقات بين المواد الجامعية | ٣٤ |
| ٧.٥ | اللغة العربية | ٣٥ |
| ٨.٥ | اللغة الإنجليزية | ٣٥ |
| ٩.٥ | الدراسات الإسلامية | ٣٦ |
| ١٠.٥ | الفيزياء | ٣٦ |
| ١١.٥ | الكيمياء | ٣٧ |
| ١٢.٥ | الأحياء | ٣٧ |
| ١٣.٥ | العلوم الهندسية | ٣٨ |
| ١٤.٥ | علوم الحاسوب | ٣٨ |
| ١٥.٥ | الفنون الجميلة | ٣٩ |
| ١٦.٥ | الرياضيات | ٣٩ |

# فهرس الجداول


| رقم الجدول | موضوع الجدول | رقم الصفحة |
|---|---|---|
| 1.2 | مثال لحساب Support | ٧ |
| 2.2 | مثال لحساب Confidence | ٧ |
| ١.٣ | جدول يوضح مقارنة بين مقترحات RARM,ARM | ١٦ |
| ٢.٣ | جدول يوضح القواعد المستخلصة بواسطة خوارزمية -Aproiri Frequent | ١٧ |
| ٣.٣ | جدول يوضح القواعد المستخلصة بواسطة خوارزمية Aproiri-Rare | ١٨ |
| ١.٥ | يوضح ملخص النتائج المتحصل عليها | ٢٨ |

# فهرس المحتويات


| الباب الأول | المقدمة | الصفحة |
|---|---|---|
| ١.١ مقدمة. | | ١ |
| ٢.١ مشكلة وأهمية البحث. | | ١ |
| ٣.١ أهداف البحث. | | ١ |
| ٤.١ منهجية البحث. | | ١ |
| ٥.١ حدود البحث. | | ٢ |
| ٦.١ هيكلية البحث. | | ٢ |
| **الباب الثاني** | **تنقيب البيانات** | |
| **الفصل الأول** | **المقدمة** | |
| ١.٢ تنقيب البيانات | | ٣ |
| 2.2 أسباب تنقيب البيانات | | ٣ |
| 3.2 خطوات إستكشاف المعرفة من قواعد البيانات | | ٤ |
| 1.3.2 إختيار البيانات(Data Selection) | | ٤ |
| 2.3.2 تهيئة البيانات(Data Preprocessing) | | ٤ |
| 3.3.2نقل البيانات(Data Transformation) | | ٥ |
| 4.3.2 تنقيب البيانات(Data Mining) | | ٤ |
| 5.3.2 تقييم النموذج(Pattern Evaluation) | | ٥ |
| 6.3.2 تمثيل المعرفة(Knowledge Representation) | | ٥ |
| 4.2 أنواع تنقيب البيانات | | ٥ |
| ١.٤.٢ التنقيب الوصفي | | ٥ |
| 2.4.2التنقيب التنبؤي | | ٦ |
| 5.2 مهام تنقيب البيانات | | ٦ |
| 1.5.2 العنقدة | | ٦ |
| 2.5.2 إكتشاف التسلسل | | ٦ |
| ٣.٥.٢ التصنيف | | ٦ |
| ٤.٥.٢ الإنحدار | | ٧ |
| ٥.٥.٢ التلخيص | | ٧ |
| قاعده الإرتباط | | ٧ |
| ٦.٢ مقاييس قواعد الإرتباط | | ٧ |
| 1.6.2 الدعم (SUPPORT) | | ٧ |
| 2.6.2 الثقة (CONFIDENCE) | | ٨ |
| 7.2خوارزمية Aproiri | | ٨ |
| 1.2 سجل الطالب | | ١٠ |
| 2.2أهمية سجل الطالب | | ١٠ |
| ٣.٢سرية السجلات | | ١٠ |
| 4.2 المعلومات الإستدلالية وسجل الطالب | | ١٠ |
| 1.4.2معلومات السجل | | ١٠ |
| 2.4.2 سجل الطالب | | ١١ |
| 3.4.2 المستندات الرسمية من السجلات | | ١١ |

| | |
|---|---|
| **الباب الثالث** **الدراسات السابقة** | |
| 1.3 استخدم تقنيه التصنيف لتقييم سجل الطالب | ١٢ |
| 2.3 إستخدم تقنيه قوانين الإرتباط | ١٣ |
| 3.3 استخدم تقنيه التنبؤ | ١٦ |
| **الباب الرابع** **منهجية البحث** | |
| 1.4 اولاً إختيار البيانات | ١٨ |
| ٢.٤ ثانيا تهيئة البيانات | ١٨ |
| ٣.٤ ثالثا تنقب البيانات | ١٩ |
| ٤.٤ ما تحتويه الويكا | ١٩ |
| 5.4 مقدمة عن الويكا | ١٩ |
| ٦.٤ صيغة ARFF | ١٩ |
| ٧.٤ طريقة إستخدام الويكا | ٢٠ |
| ٨.٤ البداية | ٢١ |
| ٩.٤ تهيئة البيانات للويكا | ٢١ |
| ١٠.٤ تحميل البيانات في المستكشف | ٢١ |
| ٤.١١ خوارزمية APROIORI | 24 |
| **الباب الخامس** **النتائج والتوصيات** | |
| ٥.١ مقدمة | ٢٦ |
| ٥.٢ مناقشة النتائج | ٢٨ |
| ٥.٣ إحصائيات البيانات | ٢٩ |
| ٥.٤ التوصيات | ٣٨ |
| ٥.٥ الخاتمة | ٣٨ |
| ٥.٦ المراجع | ٣٩ |