جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

تم إعداد هذا البحث كجزء من متطلبات نيل درجة ماجستير علوم الحاسوب

اختبار علاقة النجاح في الرياضيات في الشهادة السودانية بالنجاح في المرحلة الجامعية

An Examination of the relationship between passing mathematics in the Sudanese School Certificate and the success in university education

إعداد الطالبة:

وفاق عبدالرحيم عبدالعزيز أحمد

الإشراف:

بروفسير عزالدين محمد عثمان

ابريل 2021

الإهـــــــداء

الي امي الغالية الحنونة امد الله في عمرها الي والدي متعه الله بالصحة والعافية الى اخواني الاعزاء وإلى مشرفي البروفيسير/ عز الدين محمد عثمان الى من صبر علي وتحمل المشاق معي وبذل الجهد الكثير من اجل هذا البحث والذي مدني بدفعات معنوية زوجي ورفيق دربي الاستاذ /محمد هاشم احمد واعزهم عندي فلذة كبدي ابني /أمين متعهم الله جميعاً بالصحة والعافية.

أهدي هذا التعب والجهد لهم جميعا.

Abstract

This thesis discussed the influence of mathematic degree of the Sudanese high school Certificate result on students. The intended students are the ones who had been graduated from university at the years between 2008 until 2016 and gained math degree between 50 to 55. The reason why we choose this range of grades. Because there are no records for the student who failed in math, so this is the closest range that we can work on. This group of students had been compared to another group of student who had same percentage of certificate result but have score above 55 in math degree. The purpose of the comparison is to check the influence of the math in the students' performance in the University for Different Majorities. The analysis of data had been done using data mining algorithm named random tree algorithm. The conclusion of the thesis as mentioned below that some of Sudan university faculties students does not affected by math grade in the high school. However, to make the algorithm results more precise I recommend using more data by including other universities data.

المستخلص

ناقش هذا البحث تأثير أحد المواد الأساسية – مادة الرياضيات – في أداء الطالب في الجامعة لمختلف كليات جامعة السودان للعلوم التكنولوجيا للطلاب الذين تخرجوا في الاعوام من عام 2008م إلى عام 2016م. تم اختيار بيانات الطلاب الحاصلين على درجات تتراوح بين 50–55 في مادة الرياضيات لأنه لا توجد سجلات للراسبين في الرياضيات فتم اختيار الفئة الاقرب لاختبار النتائج وتم اختيار مجموعة اخرى لديهم نسب في الشهادة السودانية مثل نسب الفئة الاولي ولكن لديهم درجات اعلي من 55 في مادة الرياضيات. وذلك لمقارنة النتائج وتحديد تأثير مادة الرياضيات في درجات الطلاب في التعليم الجامعي.

وذلك عن طريق تحليل البيانات وتمت هذه الدراسة باستخدام خوارزمية (random tree) كخوارزمية تنقيب البيانات (data mining)، وتوصلت الدراسة إلى ان بعض الكليات جامعة السودان للعلوم والتكنولوجيا لم يتأثر أداء الطالب فيها بدرجة مادة الرياضيات حسب نتائج الطالب الأكاديمية لتلك الكليات. لزيادة دقة النتائج يجب استخدام عدد أكبر من البيانات وتضمين مجموعة من الجامعات.

الفهرس

العنوان	رقم الصفحة
الإهداء	أ
Abstract	ب
المستخلص	ج
الفهرس	د
فهرس الجداول	ه
فهرس الأشكال	ه
الباب الأول المقدمة	
1.1 المقدمة	2
2.1 مشكلة البحث	2
3.1 أهمية البحث	2
4.1 سؤال البحث	3
5.1 اهداف البحث	3
6.1 حدود البحث	3
7.1 منهجية البحث	3
8.1 هيكلية البحث	3
الباب الثاني الخلفية النظرية والدراسات السابقة	
2.1 المقدمة	6
2.2 الخلفية النظرية	6
2.2.1 تنقيب البيانات	6
2.2.2 اكتشاف المعرفة	6
2.3 أنواع تنقيب البيانات	8
2.3.1 التنقيب الوصفي	8
2.3.2 التنقيب التنبؤي	8
2.4 تقنيات تصنيف البيانات	10
2.5 التصنيف بإستخدام شجرة القرار	11
2.6 سجل الطالب	12
2.7 الدراسات السابقة	12
الباب الثالث منهجية البحث والتقنيات المستخدمة	
3.1 المقدمة	16
3.2 خطوات تنقيب البيانات	16
3.2.1 اكتشاف البيانات	16
3.2.2 المعالجة الأولية للبيانات	16
3.2.3 تكامل البيانات	17
3.2.4 اختيار البيانات	17
3.2.5 تحويل البيانات	17
3.2.6 التنقيب عن البيانات	17
3.3 منهجية البحث	22
3.3.1 الطريقة الأولى	22
3.3.2 الطريقة الثانية	23
الباب الرابع النتائج والتوصيات	
4.1 المقدمة	25

4.2 شجرة إتخاذ القرار	25
4.3 مناقشة النتائج	27
الباب الخامس التوصيات والخلاصة	
4.4 التوصيات	29
4.5 الخلاصة	29
المصادر المراجع	
الملحق	31

فهرس الجداول

الرقم	عنوان الجدول	رقم الصفحة
جدول رقم 3.1	الرياضيات أقل من 55 و الذين يوضح مقارنة بين الذين درجاتهم في مادة درجاتهم أكبر من 55	23
جدول رقم 4.1	ملخص نتائج الخوارزمية 1	25
جدول رقم 4.2	ملخص نتائج الخوارزمية 2	26
جدول رقم 4.3	يوضح تصنيف الكليات علي أساس تأثرها بدرجة مادة الرياضيات	27

فهرس الأشكال

الرقم	عنوان الشكل	رقم الصفحة
شكل رقم 1.1	يوضح خطوات اكتشاف المعرفة من قواعد البيانات	7
شكل رقم 1.2	شكل يوضح انواع تنقيب البيانات	10
شكل رقم 1.3	يوضح شجرة قرار للموافقة او رفض طلبات القروض	11
شكل 3.1	واجهة الإختيار الرسومية	19
شكل 3.2	واجهة المسكتشف في WEKA	20
شكل 3.3	طريقة فتح الملف في WEKA	21
شكل 3.4	إختيار خوارزمية التصنيف	22

الباب الأول

المقدمة

1.1 المقدمة

اتخذت وزارة التعليم العالي ضمن ثورة التعليم العالي سياسات تخص شروط قبول الطلاب للبكالوريوس في مؤسسات التعليم العالي ومن بين هذه الشروط النجاح في مادة الرياضيات وكذلك اللغة الانجليزية واللغة العربية.وقد لاحظ بعض المهتمين بالتعليم العالي انه في السابق كان هنالك بعض الطلاب الراسبين في الرياضيات قد تمكنوا من الالتحاق بالبكالريوس في التعليم العالي قبل ثورة التعليم العالي وابلوا بلاءاً حسناً في كليات الاداب والقانون وغيرهما .وقد ان الاوان لإختبار فاعلية القرار بما توفر من بيانات اداء الطلاب في الجامعات عبر العشرين سنة الاخيرة [14].

مع كثرة البيانات الموجودة والمخزنة فيما يسمي بقواعد البيانات (Database)، أصبحت موضع تساؤل عديد من الباحثين للاستفادة منها، ومع زيادة انتشار مستودعات التخزين الضخمة (data warehouses).أصبح من الضروري إيجاد تقنيات وطرق ووسائل لإستخلاص المعلومات والمعرفة من مثل هذه البيانات المكدسة واستغلالها في حل المشاكل واتخاذ القرارات. جاءت فكرة الكشف والتنقيب عن هذه البيانات بطرق ذكية للمساعدة في حل المشاكل واتخاذ القرارات.

- **تنقيب البيانات(data mining)**: هي عملية بحث محوسب ويدوي عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة[10]
- **تنقيب البيانات التعليمية(Educational Data Mining)**: هي منهجية تهتم بتطوير أساليب لاستكشاف البيانات التي تأتي من البيئات التعليمية، واستخدام تلك الأساليب من أجل فهم أفضل للطلاب و الطرق التعليمية الملائمة لهم.[11]

1.2 مشكلة البحث

اتخذت وزارة التعليم العالي ضمن ثورة التعليم العالي سياسات تخص شروط قبول الطلاب للبكالريوس في مؤسسات التعليم العالي ومن بين هذه الشروط النجاح في مادة الرياضيات وكذلك اللغة الانجليزية واللغة العربية.مما يؤدي الي مواجهة الطلاب مشكلة في الانتساب الي البكالريوس في مؤسسات التعليم العالي عند رسوبهم في هذة المواد، مع وجود احتمال تفوقهم في بعض المواد الاخرى.

1.3 اهمية البحث

تتمثل اهمية البحث في الاتي:

- الاستفادة من البيانات الضخمة في اتخاذ القرارات واختبارها.

- توفير معرفة قد تساعد وزارة التعليم العالي والبحث العلمي في اتخاذ قرارات تتعلق بقبول الطلاب في الجامعات.
- التخطيط بناءا علي التجارب بطرق علمية.

1.4 سؤال البحث

هل الرسوب في مادة الرياضيات في الشهادة السودانية يعني عدم النجاح في البكالريوس في التعليم العالي؟.

1.5 اهداف البحث

تتمثل اهداف البحث في الاتي:

- تقييم اداء الطلاب باستخدام بعض تقنيات تنقيب البيانات.
- تحليل البيانات للتوصل لمعرفة جديدة تساعد في اتخاذ قرار بخصوص الطلاب الراسبين في مادة الرياضيات .
- توفير فرص للطلاب في الانتساب لمؤسسات التعليم العالي.

1.6 حدود البحث

تتمحور حدود البحث في طلاب جامعة السودان للعلوم والتكنولوجيا الذين احرزوا اقل من 56 درجة في مادة الرياضيات في الشهادة الثانوية .

1.7 منهجية البحث

اختيار مجموعة من طلاب جامعة السودان للعلوم والتكنولوجيا كادوا ان يرسبوا في مادة الرياضيات في الشهادة السودانية ومقارنة اداءهم في بعض كليات التعليم العالي مع طلاب اخرين حاصلين علي نفس النسبة في الشهادة السودانية .تم اختيار جامعة السودان لانه يوجد بها كليات علمية وادبية (كليات تعتمد علي مادة الرياضيات كاساس للكلية واخرى لاتعتمد عليه).

1.8 هيكلية البحث

يتضمن البحث بالإضافة إلى هذا الباب:-

الباب الثاني:ويتضمن نبذة عامة عن تنقيب البيانات ،مهامها ، أنواعها بالإضافة الى الدراسات السابقة.

الباب الثالث:يتناول هذا الباب منهجية البحث.

الباب الرابع:يتناول النتائج ومناقشتها.

الباب الخامس: يتناول التوصيات والخلاصة.

يلي ذلك المصادر والمراجع والملاحق .

الباب الثاني

الخلفية النظرية

الدراسات السابقة

2.1 المقدمة

هذا الباب يتناول الخلفية النظرية لمصطلح تنقيب البيانات وكذلك يشتمل على بعض الدراسات السابقة.

2.2 الخلفية النظرية

2.2.1 تنقيب البيانات (Data Mining)

التنقيب في البيانات هو عملية تحليل كمية كبيرة من البيانات لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات، كما يمكن من الوصول إلي المعلومات التي تساعد في صناعة القرار.

يتعامل تنقيب البيانات عادة قاعدة بيانات مع بيانات يكون قد تم الحصول عمليا بغرض غير غرض التنقيب في البيانات(مثلا التعاملات في مصرف ما) مما يعني أن طريقة التنقيب في البيانات لاتؤثر مطلقاً على طريقة تجميع البيانات ذاتها. هذه هى أحد النواحي التي يختلف فيها التنقيب في البيانات عن الإحصاء، ولهذا يشار إلى عملية التنقيب في البيانات عمليا أنها عملية إحصائية ثانوية. يشير التعريف أيضا إلى أن كمية البيانات عادة كبيرة أما في حال كون كمية البيانات صغيرة فيفضل إستخدام الطرق التحليلية العادية. وتعتبر عملية تنقيب البيانات عملية تحليلية والبحث في بيانات ضخمة وهائلة لإستخراج انماط مفيده وا يجاد العلاقات ومدى الترابط بين عناصراها.[1]

2.2.2 اكتشاف المعرفة Knowledge Discovery in Database ((KDD))

اكتشاف المعرفة في قواعد البيانات ليس بالعملية السهلة والتي قد يعتقد البعض أنها تتوقف عند تجميع البيانات وإدارتها، بل نراها تمتد إلى التحليل والتوقع والتنبؤ بما سيحدث مستقبلاً.

التنقيب في البيانات يشكل جزءا من اكتشاف المعرفة(knowledge discovery)، وهذه العملية هي الأكثر شمولا[2]. تتضمن عملية اكتشاف المعرفة الخطوات التالية:

2.2.2.1 اكتشاف البيانات(Data discovery)

وهي مرحلة جمع البيانات وتشمل كشف وتحديد وتوصيف البيانات المتاحة.

2.2.2.2 تصفية البيانات وتنقيتها (Data cleaning)

ويتم في هذه المرحلة إزالة البيانات المزعجة (Noise) التي لا أهمية لها، كما يتم حذف البيانات المتضاربة والبيانات الغير متناسقة.

2.2.2.3 تكامل البيانات (Data integration)

يتم في هذه المرحلة تجميع البيانات المتشابهة وذات الصلة من مصادر البيانات المتعددة ودمجها معا.

2.2.2.4 اختيار البيانات(Data selection)

في هذه المرحلة، يتم تحديد واسترجاع البيانات الملائمة من مجموعة البيانات.

2.2.2.5 تحويل البيانات(Data transformation)

في هذه المرحلة يتم تحويل البيانات إلى نماذج مخصصة ملائمة لإجراءات البحث والاسترجاع بواسطة خلاصة الإنجاز أو عمليات التجميع.

2.2.2.6 التنقيب عن البيانات (Data mining)

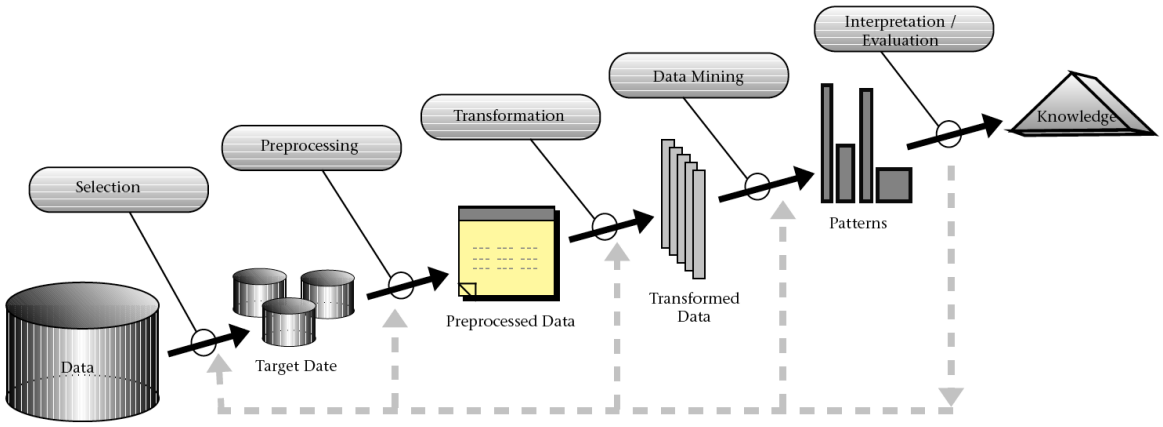
في هذه المرحلة يتم استخدام طرق ذكية تطبق لاستخلاص أنماط البيانات و استخراج نماذج مفيدة قدر الإمكان.

2.2.2.7 تقييم النمط (Pattern evaluation)

يتم في هذه المرحلة تحديد الأنماط المهمة حقا والتي تمثل قاعدة المعرفة لاستخدام بعض المقاييس المهمة.

2.2.2.8 تمثيل المعرفة وتقديمها(Knowledge presentation)

وهي المرحلة الأخيرة من مراحل اكتشاف المعرفة في قواعد البيانات وهي المرحلة التي يراها المستفيد، هذه المرحلة الأساسية تستخدم الأسلوب المرئي لمساعدة المستفيد في فهم و وتفسير نتائج استخراج البيانات.



شكل(1.1) : يوضح خطوات اكتشاف المعرفة من قواعد البيانات [2]

وبالتالي التنقيب في البيانات هو خطوة أساسية لتطبيق أساليب ذكية بهدف الكشف عن أنماط البيانات المثيرة للاهتمام والمخبأة في مجموعات البيانات الكبيرة. وهناك جانب مهم جدا، يجب النظر إليه بعين الاعتبار و هو أن هناك أنماط جديدة قد تبرز،عادة ما تكون غير معروفة من قبل. وبالتالي يجب أن تكون أدوات التنقيب عن البيانات قادرة على البحث عن أنواع مختلفة من الأنماط، بأشكال متوازية لزيادة كفاءة التنقيب عن البيانات. كما يجب أيضا أن يتم الكشف عن الأنماط في الأجزاء الصغيرة والفرعية، والتي تعرف بالحبيبات(granularities) ، مما يعني البحث في مستويات مختلفة من التجريد أو التفصيل. حلول التنقيب عن البيانات الجيدة هي التي تشير أيضا إلى قدر من الثقة أو اليقين المرتبطة مع نمط اكتشافها، لأن بعض الأنماط قد لا تصلح لكافة البيانات في مجموعة البيانات التي تم تحليلها.[2]

2.3 أنواع تنقيب البيانات

هناك نوعان أساسيان للتنقيب في البيانات هما:

2.3.1 التنقيب الوصفي:

يعتمد علي إعادة تنظيم البيانات، والتنقيب في أعماقها لإستخارج النماذج الموجوده فيها ، كتشابه الزبائن الذي يسمح لك بإنشاء وصف بسيط عن مجموعة زبائن متشابهين ولا يستوجب وجود هدف لمثل ذاك التشابه . ينتج عنه معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب. ينقسم التنقيب الوصفي الي الادوات التالية:

2.3.1.1 التجميع أو العنقدة (Clustering) :

يتمثل التجميع في البحث عن مجموعات متجانسة من العناصر .و يشير التجميع إلى عملية تشكيل مجموعات أو قطاعات ، و ذلك بالاستناد إلى معلومات متضمنة في مجموعة من المتغيرات التي تصفهم. تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

2.3.1.2 قاعدة الارتباط (Association Rules):

يتمثل الارتباط في البحث عن علاقات أو ارتباطات موجودة بين عدة خصائص.

2.3.1.3 إكتشاف التسلسل (Sequence Discovery):

ان اكتشاف التسلسل تستخدم لتحديد نماذج متسلسلة في البيانات وهذه النتائج تكون مبنية علي تسلسل الزمن في الاعمال ، وهي مماثلة للتجميع في انها تستخدم لتحديد العناصر التي تحدث معا ، ولكن الأهم من ذلك انها تستخدم لتحديد أي من العناصر يحدث أولاً[3].

2.3.2 التنقيب التنبؤي:

يحاول إيجاد افضل التنبؤات إعتماداً علي المعطيات كمعرفة المنتج الافضل لزبون معين ، بإختصار يعتمد هذا التنقيب على إستخدام معلومات قديمة لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف. ينتج عنه نموذج عن النظام الذي تصفه البيانات المستخدمة في التنقيب. ينقسم التنقيب التنبؤى الي الادوات التالية:

2.3.2.1 تحليل التسلسل الزمني (Time Series Analysis):

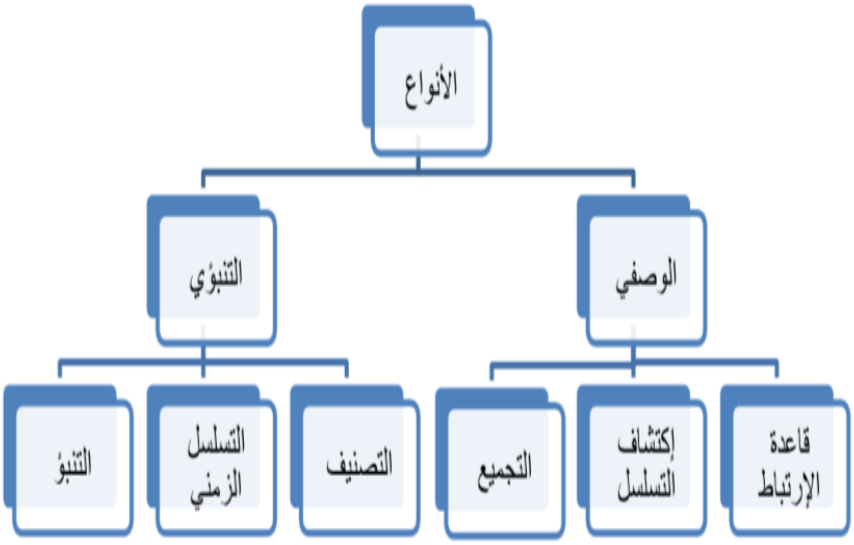
باستخدام تحليل التسلسل الزمني فان قيمة السمة يتم اختيارها كما لو انها تتغير مع الزمن[3].

2.3.2.2 التصنيف (Classification):

يتمثل التصنيف في تفسير أو التنبؤ بخاصية عنصر ما من خلال خصائص أخرى. [4]. و يمكن انجاز التصنيف بالاعتماد على الأساليب الإحصائية القديمة مثل الانحدار و التحليل التمييزي، أو بالاعتماد على أساليب حديثة نسبيا مثل قوى الارتباط و الاستنتاج المستند إلى الحالة و الشبكات العصبية.

2.3.2.3 التنبؤ (Prediction) :

يشبه التنبؤ التصنيف ، ما عدا أن البيانات تصنف على أساس التنبؤ بسلوكها المستقبلي أو تقدير قيمتها المستقبلية[5]. حيث أن المتغير التابع المتنبأ به هو متغير كمي.



شكل(1.2) : يوضح انواع تنقيب البيانات

2.4 تقنيات تصنيف البيانات (Classification Techniques)

التصنيف هو شكل من أشكال تحليل البيانات التي تستخلص نموذجا يصف فئات البيانات الهامة. ويشمل هذة النماذج الرئيسية :

2.4.1 شجرة القرار (Decision Tree) .

2.4.2 تصنيف النظرية الافتراضية (Bayesian Classification).

2.4.3 الشبكات العصبية (Neural Networks).

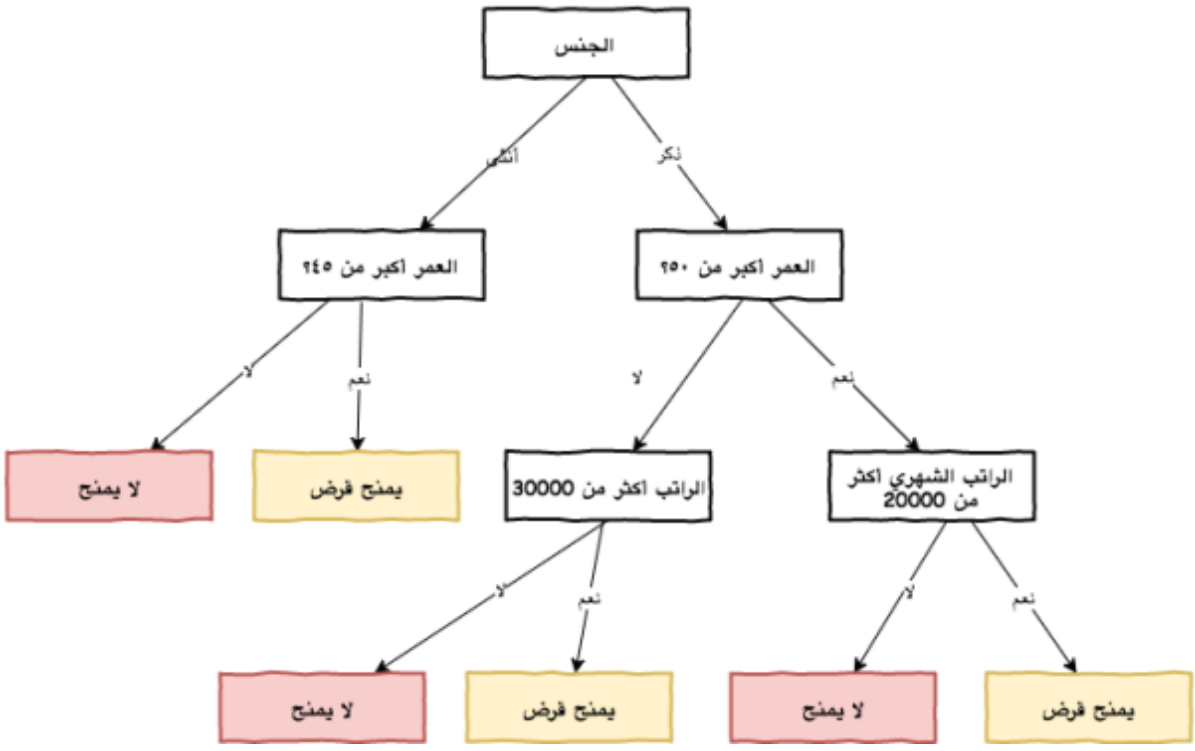
على الرغم من أن كل نموذج له سمة خاصة به، إلا أنها جميعها تشترك في خاصية معالجة البيانات في مرحلتين:

المرحلة الأولى بناء المصنف وتسمى هذه الخطوة مرحلة التعلم. ومرحلة التصنيف حيث يستخدم المصنف لتحديد فئة البيانات المعطاة.

2.5 التصنيف باستخدام شجرة القرار(Classification By Decision Tree)

تكوين شجرات القرار هو تعلم اشجار القرار من صفوف التجريب معلومة الطبقة (class labelled training set).

شجرة القرار هو مخطط انسيابي يشبه بنية الشجرة، حيث كل عقدة داخلية (عقدة غير طرفية) (internal node) تدل علي اختبار سمة(attribute) ،كل فرع يمثل نتيجة للاختبار. وكل عقدة طرفية (leaf node) تحمل تسمية فئة (class label) ، العقدة العليا في شجرة هي عقدة الجذر(root node) .



شكل(1.3) يوضح شجرة قرار للموافقة او رفض طلبات القروض

2.6 سجل الطالب

هو بيان يوضح سير الطالب الدراسي، ويشمل المقررات التي يدرسها في كل فصل دراسي برموزها وأرقامها وعدد وحداتها المقررة والتقديرات التي حصل عليها ورموز وقيم تلك التقديرات، كما يوضح السجل المعدل الفصلي والمعدل التراكمي وبيان التقدير العام بالإضافة إلى المقررات التي أعفى منها الطالب المحول. ويعرف ايضا بالسجل التربوي ويعني السجل الرسمي للطالب والذي يتم الإحتفاظ به من قبل الجامعة وتقوم بإستخدامه ويتضمن دليل المعلومات وسجل الطالب.

حفظ المعلومات وتبويبها وتنظيمها في قوالب تساعد على تيسير العمل، كما أنه يساعد علي الإستفادة من معلومات الطالب في التقييم والتطوير الأكاديمي، بالإضافة الى أنه يستخدم كوسائل توثيق يتم الرجوع إليها عند الحاجة.

2.7 الدرسات السابقة

في هذه الدراسة قام الباحث ببناء نموذج يصنف الطلاب الى خمسة فئات، اعتمادا على أدائهم الجامعي وبناء على بيانات الطالب قبل الجامعية. تقتصر الدراسة على بيانات الطلاب لثلاث سنوات جامعية (للفترة الزمنية بين(2007 –2009)عن نسبة متساوية من الذكور والإناث، مع خلفية التعليم الثانوي المختلفة من مختلف المدن والقرى البلغارية. تم تطبيق العديد من خوارزميات التصنيف المختلفة باستخدام تقنيات التصنيف المختلفة علي 10330 طالب أثناء العمل البحثي المنجز وكانت النتائج لجميع المصنفات المختبرة تؤدي أداء بدقة أقل من 70٪ مما يعني أن معدل الخطأ مرتفع والتنبؤات غير موثوقة جدا [6].

في هذه الدراسة، يتم استخدام طريقة شجرة القرار على قاعدة بيانات الطلاب تحتوي على 1548 طالب للتنبؤ بأداء الطالب. تم إستخدم بعض السمات التي تم جمعها من قاعدة بيانات الطلاب للتنبؤ بالدرجة النهائية للطالب. وسوف تساعد هذه الدراسة على تحسين أداء الطالب، وتحديد الطالب الذى يحتاج إلى اهتمام خاص للحد من الفشل ، واتخاذ الإجراءات المناسبة في الوقت المناسب [7].

هذه الدراسه تم استخدام تقنية التصنيف (Classification) والاختيار وفق الميزات (Feature selection) علي بيانات تسجيل الطالب ، لإكتشاف العوامل التي قد تؤثر على التحصيل الدراسي في

نظم المعلومات في كلية الفنون التطبيقية المفتوحة في منطقة نيوزلندا ، وكان عدد البيانات المستخدمة 453 طالب ، كانت بيانات القيد تخص المعلومات الآتيه:

الجنس ، تاريخ الميلاد، الأصول العرقية والعجز، حالة العمل ، تأهيل المدارس الثانويه الأكاديمية، Programming course، Collage course، course block، تفاصيل الإتصال.يتم ملئ إستمارةالتسجيل في integrator يتم تسجيل تاريخ التسجيل ويصبح الطالب مسجل في (course).

وكانت تقنيات تنقيب البيانات المستخدمة هي التصنيف وإختيار ميزة(feature selection) ، وتم إستخدام أربع أنواع من أشجار تصنيف وهي:

- CHAID.
- Exhaustive CHAID.
- QUEST.
- CART.

من خلال نتائج البحث وجد أن CARTأوجدت العوامل المهمة التي تفصل بين الطلاب الناجحين والفاشلين وهي:

- الاصول العرقية.
- Programming Course.
- Course block.

ولكن لسوء الحظ أشجار التصنيف لم تعطى دقة عالية في النتائج ، وكان دقة شجرةCHAID هي 59.4% وكانت دقة شجرة CART هي 60.5% ، وهي تقترح أن المعلومات المساعدة (الجنس،العُمر ، إلإنتماءالعرقي ، العجز ، المدرسةالثانوية ، حالة العملِ ، والتسجيل المبكر) تجمع أثناء عملية التسجيلِ ، ولاتحتوي معلومات كافية للفصل بين الناجحين والفاشلين.تهتم هذه الدراسة بالنواحي الإجتماعية والبيئة الدراسية التي تؤثر علي الطالب ، بمعنى اخر من خلال بيانات التسجيل نستطيع التحكم في مدى نجاح او فشل الطالب [8].

قام الباحثين بدراسة في احدى جامعات الهند بعمل بحث بإستخدام تقنيه التصنيف وتم تطبيقها علي البيانات التعليمية أخذت من درجات طلاب الماجستير وكان عدد الطلاب 60 طالب ، وكانت البيانات

تتضمن رقم الطالب ومجموع العلامات التي تحصل عليها خلال الأربعه سنوات الدراسية، وكان الغرض من هذا البحث هو توقع أداء الطالب في الإمتحانات وإظهار التحسن في أداءه الأكاديمي.

وتم إعداد البيانات(Preprocessing) علي مجموعة البيانات طبقا لقواعد الجامعة والعلامات التي حصل عليها الطالب تم تحويلها إلى الشكل الإسمي(nominal) وذلك لغرض البحث ، وتم إستخدام إحدى أدوات تنقيب البيانات وهى (weka).وكانت العوامل المستخدمة في الخوازميتين هي:

- الكفاءة في توقع نتيجة الطالب.
- القيم الايجابية الصحيحة والقيم الايجابية الخاطئة وقيم ال ROC.
- الزمن المأخوذ في كل خوارزمية لإنتاج الناتج.

تقنيات التنصيف المختلفة يمكن إستخدامها لتوقع نتيجه الطالب في الفصل الدراسي النهائي بحيث يكون مستند علي العلامات التي حصل عليها في الفصول الدراسية السابقة ، وفي هذا البحث يتم مراقبة البيانات المتوفرة في القسم وإستخدامها في توقع نتيجة الطالب في الفصل الدراسي القادم.

تقنيات التصنيف التي تم تطبيقها هي:

- One-R Algorithm.
- Decision Table algorithm.

بعد تطبيق التقنيتين وجد أن خوارزمية (One-Rule)أفضل من خوارزمية (Decision Table)لعدة اسباب وهي أن الخوارزمية الأولى أكثر دقة من الأخرى وتؤدي أفضل منها وتأخذ وقت اقل[9].

الباب الثالث

منهجية البحث والتقنيات المستخدمة

3.1 المقدمة

بالتعامل مع سجل الطالب الأكاديمي نجد الكثير من المعلومات الأكاديمية المتعلقة بالطالب ، وفي هذا البحث تم تحديد بعض هذه البيانات والتعامل معها وقد تمت معالجتها وفقا لخطوات تنقيب البيانات.

3.2 خطوات تنقيب البيانات:

1.2.3 اكتشاف البيانات (Data discovery)

تم الحصول علي البيانات علي قسمين :

- **القسم الاول :**

من وزارة التعليم العالي وكانت البيانات لطلاب جامعة السودان للعلوم والتكنولوجيا تم اختيار مجموعة من الصفات وهي درجات الطالب في مادة الرياضيات(متمثلة في الرياضيات الاساسية والرياضيات المتخصصة) في الشهادة الثانوية ، رقم الجلوس، النوع ، رقم الاستمارة ، سنة القبول، نسبة النجاح والكلية التي تم قبول الطالب بها.

- **القسم الثاني:**

من قسم مركز البيانات بجامعة السودان للعلوم والتكنولوجيا للطلاب الذين تخرجوا في الاعوام من عام 2008م إلى عام 2016م وتتمثل في الرقم الجامعي للطالب، النوع ، رقم الكلية ،المعدل الفصلي ، المعدل التراكمي ،تفاصيل النتيجة و سنة الدخول.

2.2.3 المعالجة الاولية للبيانات(Data Pre-Processing)

مثلت البيانات العمود الفقري لعملية التنقيب للإكتشاف المعرفة والبيانات التي تم التعامل معها لم تكون في الشكل الملائم لعملية التنقيب .فالتحدي الصعب كان هو تجهيزها في الشكل الملائم لعملية التنقيب.

وقد أخذ 70% تقريباً من زمن عملية التنقيب في تجميع وتجهيز البيانات وقد كان التجميع من ،مصادر مختلفة لذا فان تجهيز البيانات أخذ جهد و وقت حتى تم إخراج النتائج . البيانات التى تم جمعها كانت غير مكتملة ، ومشوشة ومتناقضة مما قد يخفي أنماط مفيدة ، غير مكتملة بسبب عدم توفر بعض القيم لبعض الخواص والمتغيرات بل كانت هنالك متغيرات غير موجودة أصلاً.

وتم إختيار البيانات التي لها علاقة مع بعضها البعض وذلك عن طريق اختيار بعض الخواص وعمل ترشيح لبعضها حيث تم مسح البيانات الشاذة و إزالة السجلات المتكررة .

بعد إكتمال جمع البيانات من مصادرها واستبعاد البيانات الغير مكتملة البيانات من مجموعة البيانات لأن هذه البيانات لابد ان تكون مكتملة لأنها تحتوي على درجات طلاب وهذه الدرجات لايمكن تعويضها بشئ اخر لانها لابد ان تكون حقيقة.

3.2.3 تكامل البيانات (Data Integration)

في هذه المرحلة كان لابد من وجود مفتاح (Primary Key) للربط بين البيانات وكانت هذه مشكلة حقيقية حيث ان الرقم الجامعي للطالب هو المفتاح للبيانات من جامعة السودان و (رقم الجلوس + سنة الامتحان)هي المفتاح للبيانات من التعليم العالي وهي قيم غير متطابقة تم تطبيق خوارزمية لايجاد مفتاح مشترك عن طريق اضافة سنة الدخول + رقم الاستمارة لينتج عنها الرقم الجامعي للطالب حتي يتم جمع البيانات مع بعضها البعض.

4.2.3 اختيار البيانات (Data Selection)

تم اختيار بيانات الطلاب الحاصلين علي درجات تتراوح بين 50-55 في مادة الرياضيات لانه لاتوجد سجلات للراسبين في الرياضيات فتم اختيار الفئة الاقرب لاختبار النتائج وتم اختيار مجموعة اخرى لديهم نسب مثل مدي نسب النجاح في الشهادة السودانية لمجموعة الطلاب ضعيفي نسبة الرياضيات ولكن لديهم درجات اعلي في مادة الرياضيات لمقارنة النتائج وتحديد تاثير مادة الرياضيات في درجات الطلاب في التعليم العالي.

5.2.3 تحويل البيانات(Data transformation)

تم تحويل البيانات الى صيغة (Excel sheet) بالإضافة الي تهيئة الشكل العام للملف بحيث يناسب أداة تنقيب البيانات المستخدمة وهي (WEKA)وهي تقبل البيانات في صورة ملف (CSV).

6.2.3 التنقيب عن البيانات (Data mining)

في هذه المرحلة تم إستخدام تقنية التصنيف بإستخدام برنامج (Weka).

1.6.2.3 WEKA

Waikato Environment for knowledge Analysis بيئة وايكاتو لتحليل المعرفة وهي احد فروع الذكاء الاصطناعي حزمة مشهرة من تطبيقات التعليم الالى مكتوبة بلغة جافا ، تم تطويرها في العمومية جامعة وايكاتو في نيوزيلندا. WEKA من البرمجيات المفتوحة المصدر بموجب رخصة جنو العمومية العامة [10].

2.6.2.3 طريقة إستخدام(Weka)

أسهل الطرق لإستخدام(Weka) هو من خلال إستخدام واجهة المستخدم الرسومية وتسمى المستكشف (Explorer) ، وهذا الأخير يتيح الوصول الى جميع منشأت (Weka)بإستخدام إختيار القائمة وملأ الإستمارة.مثلا يمكن بسرعة كبيرة قراءة قاعدة البيانات من ملف (CSV) وتكوين شجرة القرارات منها.لكن شجرة القرارات تعتبر فقط البداية لأنه يوجد هنالك الكثير من الخوارزميات لابد من إكتشافها ،واجهة المستكشف تساعدك على فعل هذا ، يمكنها توجيهك بعرض الخيارات في شكل قوائم.

هنالك نوعين آخرين من واجهات المستخدم (Weka) بالإضافة للمستكشف، واجهة تدفق المعلومات (Knowledge Flow) وهي تسمح بتصميم الترتيبات لمعالجة البيانات المجرية(Streamed data) والواجهة الثالثة (Weka)هي المجرب (Experimental) وهي مصممة للإجابة على الأسئلة الأساسية للعملية عند الطلب على تقنيات الإرتداد والتصنيف .بالإضافة الى هذه الواجهات التفاعلية توجد الوظيفة الأساسية لل(Weka) وهي التي يمكن الوصول اليها في شكل صف بإدخال الأوامر النصية ، وهي بدورها تسمح بالوصول لكل المميزات بالنظام.عندما يعمل (Weka) لابد من الإختيار من بين هذه الأربع واجهات.

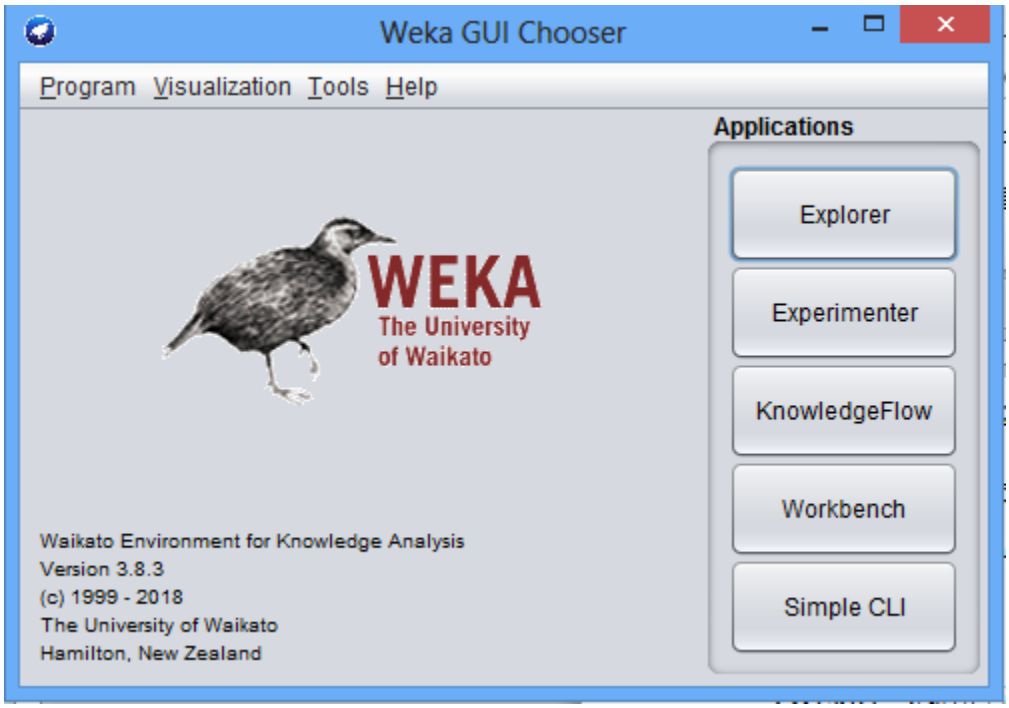
افترض أن لديك مجموعة من البيانات وتريد بناء شجرة القرارات منها ، أولا لابد من تهيئة هذه البيانات وإطلاق المستشكف والقيام بتحميل البيانات فيه ، ثانيا لابد من إختيار طريقة بناء شجرة القرار و بناء الشجرة وتفسير المخرجات.في المستشكف يمكنك التنقل ذهاباً وإياباً بين النتائج التي تحصلت عليها ، تقييم النماذج التي تم بناءها على مختلف قواعد البيانات ، تصور بشكل تخطيطي كلتا النماذج وقواعد البيانات انفسهم والتي تتضمن أخطاء التجميع التي يمكن أن يصنعها النموذج.

3.6.2.3 تهيئة البيانات(Weka)

غالبا ما يتم عرض البيانات في صورة (Excel Sheet)أو قاعدة بيانات ، على أي حال الطريقة الأصلية للتخزين في (Weka)هي في صورة (ARFF)ويمكن بكل سهولة التحويل من صورة برنامج الجدولة الى صورة (ARFF).معظم ملفات (ARFF)تتكون من قائمة من الحالات وقيم خواص لكل من الحالات التي تم الفصل بينها بالفاصلة كما ذكرنا سابقا.معظم برامج الجدولة وبرامج قواعد البيانات تسمح بتصدير البيانات في ملف (Comma Separated Value CSV)في شكل مجموعة من السجلات مع وجود الفاصلة بين القيم .

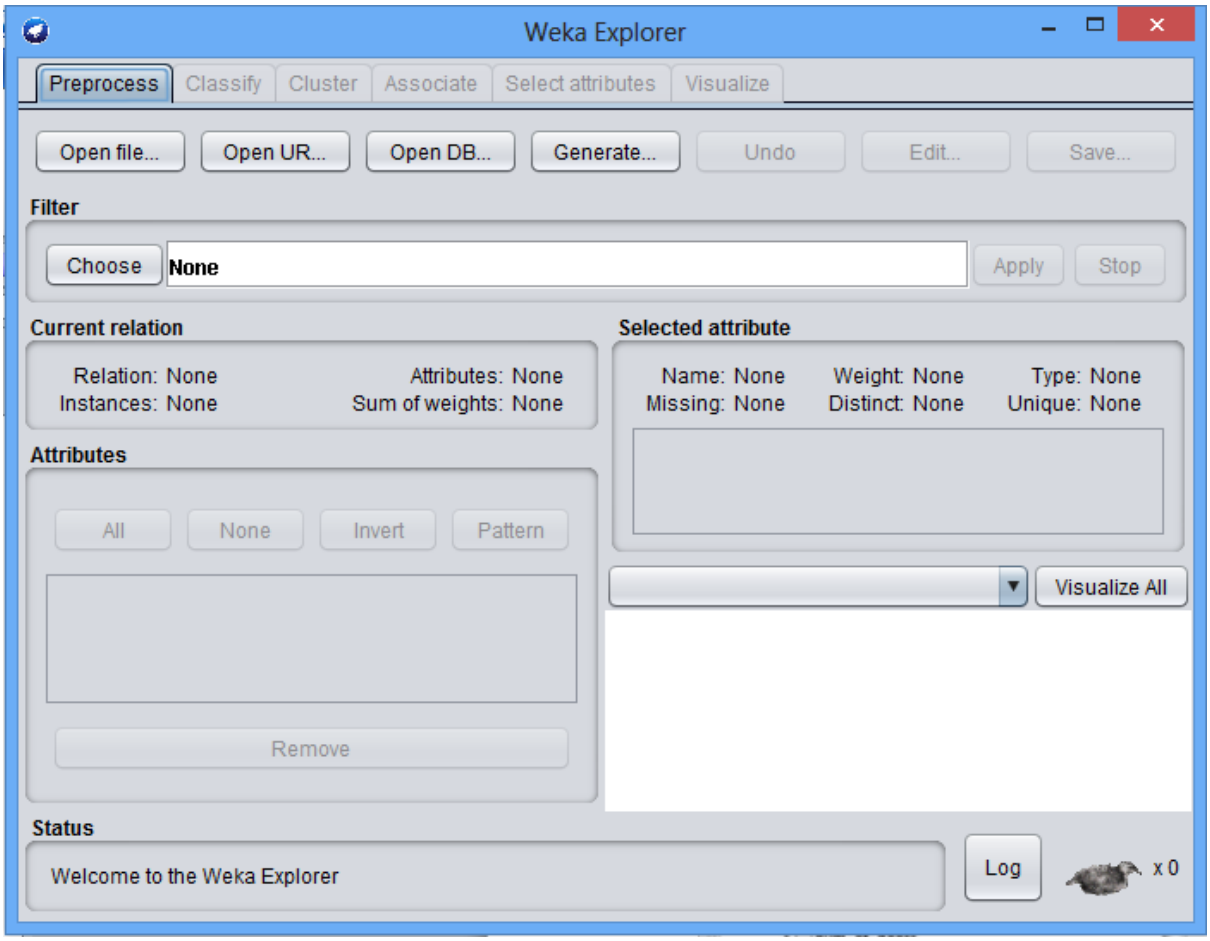
4.6.2.3 تحميل البيانات في المستكشف(Explorer)

لنبدأ تحميل البيانات في المستكشف والبداية في تحليلها ، في البدء نقوم بتشغيل برنامج (Weka) للحصول على واجهة الإختيار الرسومية كما هو موضح بالشكل التالي :



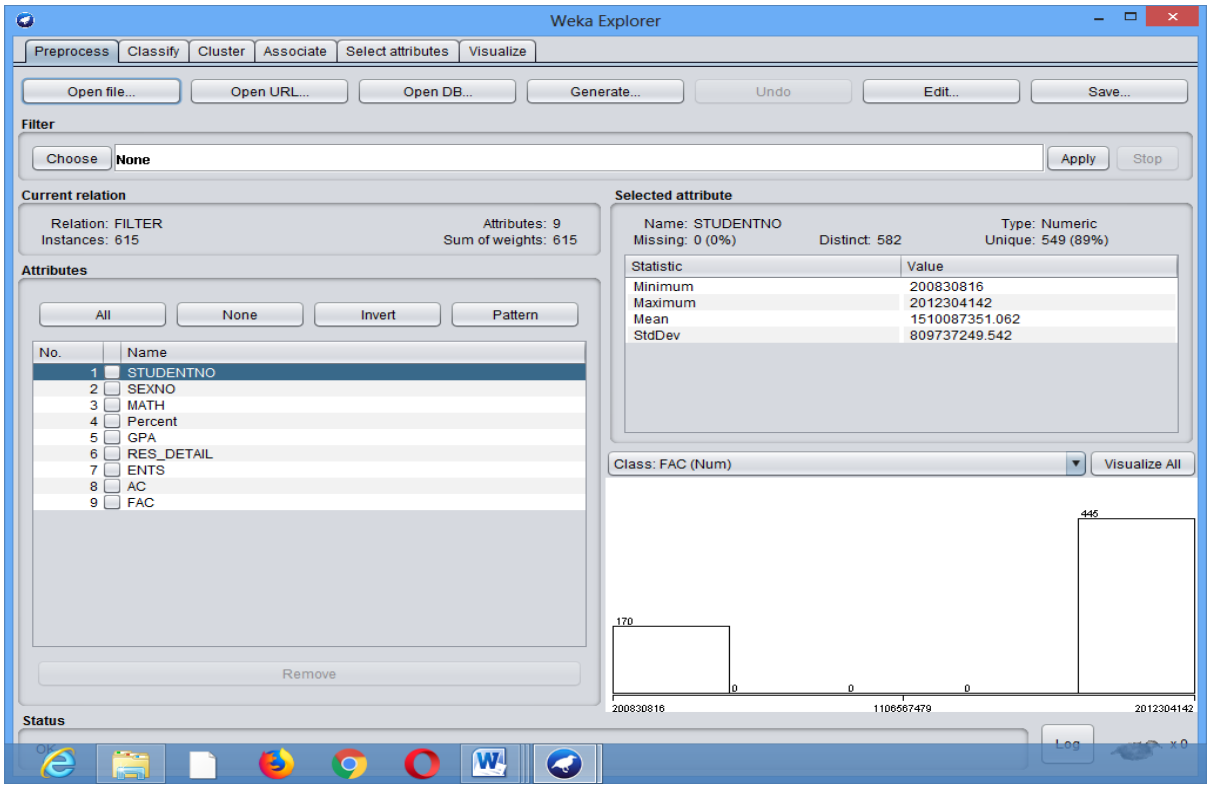
الشكل (1.3) : واجهة الاختيار الرسومية.

ثم نقوم بإختيار المستكشف من الأربع خيارات المتاحة على الجهة اليمنى من الشكل.

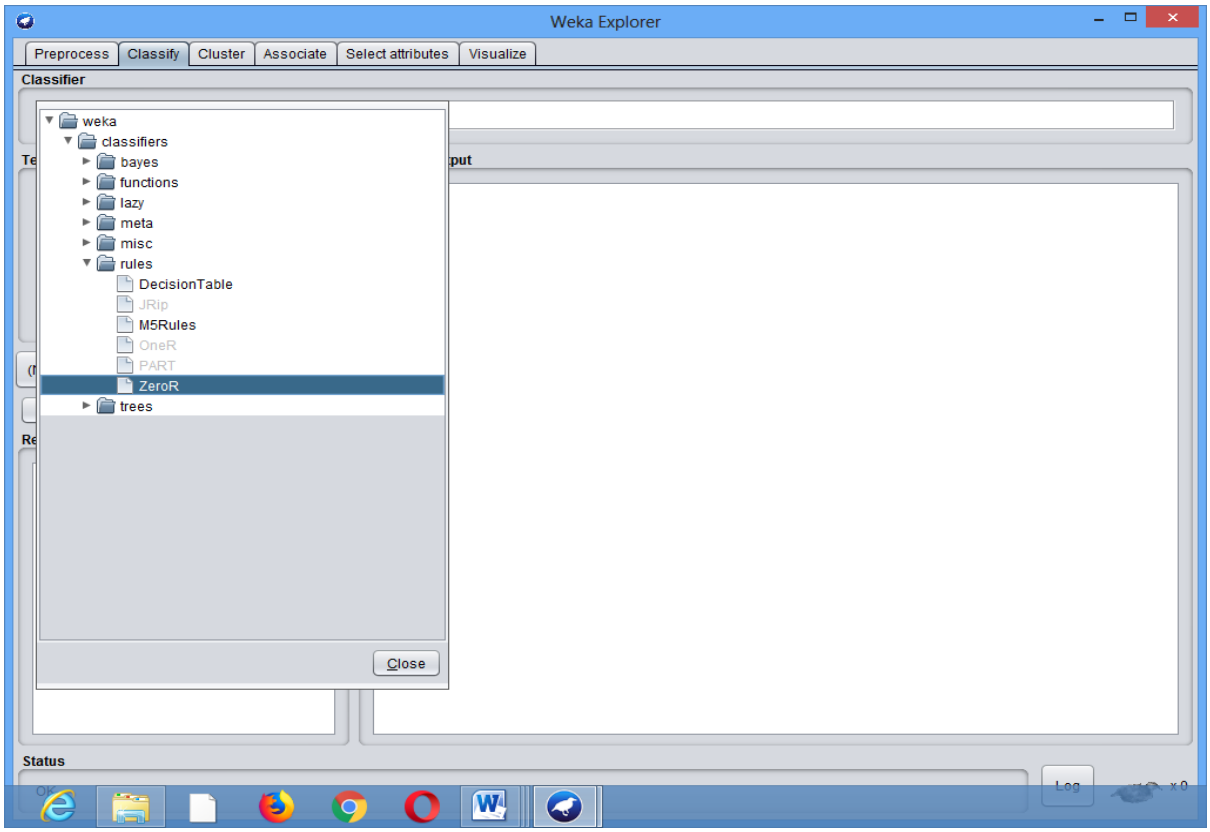


الشكل (2.3) : واجهة المستكشف في weka.

في الواقع الشكل يوضح ما يتم عرضه بعد القيام بتحميل البيانات.الست قوائم الأولى هي عبارة عن العمليات الأساسية التي يدعمها المستكشف ، حاليا نحن في مرحلة المعالجة المبدئية ، اضغط زر "افتح ملف" الذي من خلاله يمكن ان تختار الملف الذي تريد تحميله . إختار الملف ذو الامتداد (csv).بعد ان يتم تحميل البيانات نقوم بتحديد الصفات(attributes) التي نريد إجراء عملية التصنيف عليها ويوجد أيضا خيار إختيارها جميعها وإجراء التصنيف عليها كما في الشكل(3.3). ومن ثم بالرجوع للست قوائم الأولى نقوم بإختيار classify ثم نختار خوارزمية التصنيف كما في الشكل (4.3)ثم نقوم باختيار start ، بعدها يقوم البرنامج بعمل توليد للنتائج والمخرجات كما هو موضح في الشكل (5.3) :



الشكل(3.3) يوضح طريقة فتح الملف



الشكل(4.3) يوضح اختيار خوارزمية التصنيف

3.3 منهجية البحث :

تم مقارنة البيانات في البحث باستخدام طريقتين كالاتي:

3.3.1 الطريقة الاولى:

تم استخدام بيانات مكونة من 1648طالب بعد ان تم حذف الطلاب الذين هم خارج مدى نسب الطلاب الذين درجاتهم اقل من او يساوي 55 في كل كلية علي حدة . تم تقسيم الطلاب إلى مجموعات علي حسب الكلية وكانت المقارنة علي اساس نسبة الشهادة السودانية حيث تم ايجاد اكبر نسبة واقل نسبة في الشهادة السوادنية للطلاب الذين كانت درجاتهم في مادة الرياضيات بين 50 و55 الذين إلتحقوا بكلية معينة وتم مقارنتها بالطلاب الذين لديهم نسب في هذا المدى وكانت درجاتهم في مادة الرياضيات أكبر من 55. ونذكر كمثال كلية الدرسات الزراعية عدد الطلاب 993 طالب الذين درجاتهم في مادة الرياضيات اقل من او يساوي 55 هو 95 طالب 46 منهم لم يكملوا دراستهم لأسباب مختلفة وعدد الطلاب الذين درجاتهم في مادة الرياضيات اكبر

من 55 هو 898 منهم 329 لم يكملوا دراستهم بعد. تم حذف276 طالب هم الطلاب الذين درجاتهم خارج مدي نسب الطلاب الذين درجاتهم اقل من او يساوي 55 اصبح العدد الكلي للطلاب 717 (عدد الطلاب الذين درجاتهم في مادة الرياضيات اكبر من 55 هو622 طالب . وعدد الطلاب الذين درجاتهم في مادة الرياضيات اقل من او يساوي 55 هو 95 طالب).

3.3.2 الطريقة الثانية :

تم استخدام بيانات مكونة من 2447 طالب تم تقسيم الطلاب إلى مجموعتين المجموعة الاولي تحوي الطلاب الذين كانت درجاتهم في مادة الرياضيات اقل من 55 تم ايجاد مدي النسب في الشهادة السودانية وكانت بين 50 الى 74.14 ثم تمت مقارنة اداء الطلاب الذين لديهم نفس نسبة الشهاة السودانية في المجموعتين وتم مقارنة اداءهم علي اساس تقديرهم الجامعي.

وجه المقارنة	الطلاب الذين درجاتهم في مادة الرياضيات أقل من 55	الطلاب الذين درجاتهم في مادة الرياضيات أكبر من 55
عدد المتخرجين	134	1328
عدد الذين لم يكملوا دراستهم	122	863
العدد الكلي	256	2191
اكبر نسبة	74.14	74.14
اقل نسبة	50	51.14
نسبة المتخرجين	%52.34	%60.61
نسبة الذين لم يكملوا دراستهم	%47.66	%39.39

جدول(3.1) يوضح مقارنة بين الذين درجاتهم في مادة الرياضيات أقل من 55 و الذين درجاتهم أكبر من 55

الباب الرابع

النتائج

4.1 المقدمة

يهدف البحث لتحليل البيانات للتوصل لمعرفة جديدة تساعد في اتخاذ قرار بخصوص الطلاب الراسبين في مادة الرياضيات ، تم تطبيق خوارزمية (random tree)للوصول الي الهدف ، تم استخدام مستكشف الويكا (weka explorer) في هذه المرحلة.

4.2 شجرة اتخاذ القرار

تم تطبيق خوارزمية(J48) وكانت البيانات المصنفة بصورة صحيحة بنسبة 50.88% و خوارزمية (Bayes Naive) وكانت البيانات المصنفة بصورة صحيحة بنسبة 53.25% وتم تطبيق خوارزمية (Random Tree)علي البيانات وقد حققت أعلى دقة لذلك تم اختيارها. وكانت البيانات المصنفة بصورة صحيحة بنسبة 88.90%عند المقارنة باستخدام الكليات كما هو موضح في الشكل (1.4) و 94.56% عند المقارنة باستخدام نسبة الشهادة السودانية كما هو موضح في الشكل (2.4).

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.005	0.961	1.000	0.980	0.978	1.000	0.999	in comp
	0.983	0.149	0.851	0.983	0.912	0.833	0.986	0.978	Good
	0.825	0.025	0.908	0.825	0.865	0.828	0.989	0.955	Fine
	0.730	0.008	0.944	0.730	0.823	0.803	0.990	0.934	Very Good
	0.462	0.000	1.000	0.462	0.632	0.673	0.994	0.815	Excellent
Weighted Avg.	0.889	0.077	0.896	0.889	0.884	0.837	0.989	0.962	

جدول رقم(1.4) يوضح ملخص نتائج خوارزمية 1 (Random Tree)

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.998	0.044	0.938	0.998	0.967	0.945	0.999	0.997	in comp
	0.950	0.036	0.924	0.950	0.937	0.908	0.996	0.988	Good
	0.874	0.003	0.981	0.874	0.925	0.915	0.998	0.984	Fine
	0.855	0.001	0.992	0.855	0.918	0.912	0.998	0.982	Very Good
	0.877	0.000	1.000	0.877	0.935	0.935	1.000	0.988	Excellent
Weighted Avg.	0.946	0.030	0.948	0.946	0.945	0.925	0.998	0.990	

جدول رقم(2.4) يوضح ملخص نتائج خوارزمية 2 (Random Tree)

1.2.4 مصطلحات جدول النتائج:

- **الإستدعاء (Recall)أو المعدل الإيجابي الحقيقي ((True positive rate (TP Rate)):**

معدل الحالات التي صنفت بصورة صحيحة لفئة معينة. وتحسب كما يلي :

.R = TP / (TP + FN)

- **المعدل الإيجابي الخاطئ ((False positive rate (FP Rate)):**

معدل الحالات التي صنفت بشكل غير صحيح لفئة معينة، يتم الحكم عليه كعينة إيجابية، لكنه في الواقع عينة سلبية. يسمى أيضًا بالإيجابية الزائفة. وتحسب كما يلي:

FP=FN / TP+FN

- **الدقة(Precision):**

معدل الحالات الصحيحة لفئة معينة، وتنبئها المصنف بشكل صحيح و تحسب كما يلي:

P = TP/ (TP+FP)

- **مقياسF (F-Measure):**

هو مقياس دقة الاختبار، يتم حسابه من دقة الاختبار والإستدعاء و تحسب كما يلي:

F = 2 * P * R / (P + R)

- **منحنى خصائص تشغيل المستقبِل((Receiver Operating Characteristics(ROC)) :**

واحدة من أهم القيم الناتجة من Weka. يعطي فكرة عن كيفية أداء المصنفات بشكل عام.

4.3 مناقشة النتائج

بعد دراسة النتائج المتحصل عليها تم الوصول للنتائج الاتية :

الكليات التي لم تؤثر درجة الرياضيات فيها على أداء الطالب الجامعي	الكليات التي أثرت فيها درجة الرياضيات على اداء الطالب الجامعي
كلية الدراسات الزراعية	كلية علوم الغابات والمراعى
كلية علوم وتكنولوجيا الإنتاج الحيواني	كلية التربية الفنية
كلية التربية التقنية	كلية التربية البدنية والرياضة
كلية الفنون الجميلة والتطبيقية	كلية هندسة وتكنولوجيا الصناعات
كلية علوم الأشعة الطبية	كلية علوم المختبرات الطبية
كلية الهندسة قسم الهندسة الميكانيكية وقسم الهندسة الطبية وقسم الهندسة الكهربائية.	كلية الهندسة قسم الهندسة المدنية وهندسة المساحة وهندسة الطيران
كلية العمارة والتخطيط	كلية علوم الحاسوب وتقانة المعلومات
كلية الدراسات التجارية	كلية العلوم
كلية هندسة المياه والبيئة	كلية الموسيقا والدراما

جدول(4.3) يوضح تصنيف الكليات علي أساس تأثرها بدرجة مادة الرياضيات

الباب الخامس

التوصيات والخلاصة

1.5 التوصيات

1. لإكمال الأبحاث في هذا المجال يجب توفير البيانات وتخزينها بصورة أفضل بحيث تكون في متناول الجميع.
2. استخدام بيانات من جامعات نسب القبول فيها منخفضة.
3. توسيع دائرة تطبيق الخوارزمية وذلك بتطبيقها على جامعات أكثر ومعرفة النتائج في تلك الجامعات ومقارنة الجامعات ببعضها البعض.
4. استخدام بيانات أكثر لزيادة دقة النتائج.
5. استخدام الولاية كأحد خواص التصنيف لانها من العوامل المؤثرة في درجات الطلاب.
6. زيادة بيانات طلاب الكليات الأدبية.
7. استخدام بيانات تشمل جميع الكليات للحصول على نتائج أكثر دقة.
8. تطبيق التصنيف بإضافة خواص اخري لها تاثير علي درجات الطلاب.

2.5 الخلاصة

في الختام وبعد الإطلاع على النتائج وتحليها ظهرت اهمية سجل الطالب الأكاديمي في المرحلة الثانوية ومدى تأثير مادة الرياضيات على مستواه الأكاديمي في الجامعة وبذلك تظهر ضرورة وحوجة تحليل البيانات للمساعدة في إتخاذ القرارات بصورة فعالة. وكذلك يشير البحث إلى ان تنقيب البيانات من الطرق الحديثة والفعالة جداً في تجميع وتحليل المعلومات والبيانات والتوصل إلى نتائج أفضل تساعد في صنع قرارات صائبة في كثير من الأحيان.

تم الحصول علي البيانات من وزارة التعليم العالي و جامعة السودان للعلوم والتكنولوجيا للطلاب الذين تخرجوا في الاعوام من عام 2008م إلى عام 2016م. و تم إختيار البيانات التي لها علاقة مع بعضها البعض و تم مسح البيانات الشاذة و إزالة السجلات المتكررة . بعد اكتمال جمع البيانات من مصادرها واستبعاد البيانات الغير مكتملة البيانات من مجموعة البيانات. تم اختيار بيانات الطلاب الحاصلين على درجات تتراوح بين 50–55 في مادة الرياضيات لأنه لا توجد سجلات للراسبين في الرياضيات فتم اختيار الفئة الاقرب لاختبار النتائج وتم اختيار مجموعة اخرى لديهم نسب مثل مدي نسب النجاح في الشهادة السودانية لمجموعة الطلاب ضعيفي نسبة الرياضيات ولكن لديهم درجات اعلي في مادة الرياضيات لمقارنة النتائج وتحديد تأثير مادة الرياضيات في درجات الطلاب في التعليم العالي.

وتم تطبيق خوارزمية (Random Tree)على البيانات وقد حققت أعلى دقة لذلك تم اختيارها. وكانت البيانات المصنفة بصورة صحيحة بنسبة 88.90%عند المقارنة باستخدام الكليات و94.56% عند المقارنة باستخدام نسبة الشهادة السودانية.

بعد دراسة النتائج المتحصل عليها تم الوصول ان:

- الكليات الاتية لم تؤثر درجة الرياضيات فيها على أداء الطالب الجامعي:

كلية الدراسات الزراعية ، كلية علوم وتكنولوجيا الإنتاج الحيواني ، كلية التربية التقنية ، كلية الفنون الجميلة والتطبيقية ، كلية علوم الأشعة الطبية ، كلية الهندسة قسم (الهندسة الميكانيكية ، الهندسة الطبية و قسم الهندسة الكهربائية) ، كلية العمارة والتخطيط ، كلية الدراسات التجارية وكلية هندسة المياه والبيئة.

- الكليات الاتية التي أثرت فيها درجة الرياضيات على اداء الطالب الجامعي:

كلية علوم الغابات والمراعى ، كلية التربية الفنية ، كلية التربية البدنية والرياضة ، كلية هندسة وتكنولوجيا الصناعات ، كلية علوم المختبرات الطبية ، كلية الهندسة قسم (الهندسة المدنية ، هندسة المساحة وهندسة الطيران) ، كلية علوم الحاسوب وتقانة المعلومات و كلية العلوم و كلية الموسيقا والدراما.

بعد الإطلاع على النتائج وتحليلها تم ايجاد ان درجة الرياضيات قد يكون شرط غير مؤثر في النجاح لبعض الكليات في المرحلة الجامعية.

المصادر والمراجع

[1] H.A. Edelstein, Introduction to data mining and knowledge discovery, 3rd d, Potomac, MD: Two Crows Corp, 1999.

[2] Cipolla, Emil T., "Data Mining: Techniques to Gain Insight Into Your Data," Enterprise Systems Journal, Dec. 1995, vol. 10, No. 13, pp. 18 (5)..

[3] خضر مصباح و إسماعيل طيطي ، في إدارة المعرفة التقنيات والتحديات والحلول, عمان, دار الحامد للنشر والتوزيع, 2009.

[4] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Procedia Computer Science*, vol. 57, pp. 500–508, Mar 2015.

[5] G. S. Linoff and M. J. A. Berry, in Data Mining Techniques: For marketing, sales, and Customer Relationship Management, 3rd ed., Indianapolis, IN: Wiley Publishing, 2011.

[6] D. Kabakchieva, "Predicting student performance by using data mining methods for classification," *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 61–72, Feb 2013.

[7] A. B. Ahmed and I. S. Elaraby, "Data Mining: A prediction for student's performance using classification method," *World Journal of Computer Application and Technology*, vol. 2, no. 2, pp. 43–47, 2014.

[8] Z. J. Kovacic, "Early prediction of Student Success: Mining Students Enrolment Data," *Proceedings of the 2010 InSITE Conference*, 2010.

[9] S. Anupama Kumar and M. N. Vijayalakshmi, "Mining of Student Academic Evaluation Records in higher education," *2012 International Conference on Recent Advances in Computing and Software Systems*, Apr 2012.

[10] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam: Morgan Kaufmann Publishers, 2012.

[11] S. Rai and A. Kumar Jain, "Students' dropout risk assessment in undergraduate courses of ICT at Residential University A Case Study," *International Journal of Computer Applications*, vol. 84, no. 14, pp. 31–36, 2013.

[12V. Ramesh, P. Parkavi, and K. Ramar, “Predicting student performance: A statistical and data mining approach,” *International Journal of Computer Applications*, vol. 63, no. 8, pp. 35–39, 2013.

[13] B.K. Bharadwaj and S. Pal, “Data Mining: A prediction for performance improvement using classification”, *International Journal of Computer Science and Information Security* (IJCSIS), Vol. 9, No. 4, pp. 136-140, Apr 2011.

ملحق (1): يوضح البيانات المستخدمة في المقارنة باستخدام الكليات

MATH	Percent	RES_DETAIL	FACNAME	Name code
52	65.29	in comp	كلية علوم الغابات والمراعى – علوم الغابـات	FO–SC–45
55	63	in comp	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG–52
53	64.14	in comp	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG–52
55	63.14	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
54	66.85	in comp	كلية الدراسات الزراعية – الاقتصاد الزراعي	AG–29
53	64.85	in comp	كلية علوم وتكنولوجيا الإنتاج الحيواني – علوم الأسماك والحياة البرية	ANI–31
53	62.43	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
53	67.85	in comp	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI–40
54	67.29	in comp	كلية الدراسات الزراعية – الهندسة الزراعيـة	AG–80
55	63.57	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
51	63.14	in comp	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG–52
55	67.43	in comp	كلية التربية الفنية	ED–AR–36
55	65.70	in comp	كلية الدراسات الزراعية – البســاتين	AG–38
55	63.85	in comp	كلية التربية التقنية – مدنية	ED–TE–25
54	62.14	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
55	64.56	in comp	كلية الدراسات الزراعية – البســاتين	AG–38
54	66.43	in comp	كلية الدراسات الزراعية – الهندسة الزراعيـة	AG–80
54	62.14	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
55FF	63.71	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
51	64	in comp	كلية الدراسات الزراعية – البســاتين	AG–38
53	64.29	in comp	كلية علوم الغابات والمراعى – علوم المراعى	FO–SC–39
54	69.70	in comp	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI–40
55	66.29	in comp	كلية الدراسات الزراعية – الاقتصاد الزراعي	AG–29
53	66.14	in comp	كلية التربية الفنية	ED–AR–36

72	68.14	Very Good	كلية التربية الفنية	ED-AR-36
62	68.14	Fine	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG-52
79	68.14	Good	كلية التربية الفنية	ED-AR-36
68	68.14	Very Good	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI-40
60	68.14	Fine	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
68	68.14	Very Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
76	68.14	Good	كلية الفنون الجميلة والتطبيقية	AR-63
70	68.14	Fine	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI-40
63	68.14	Good	كلية علوم الغابات والمراعى – علوم الغابـات	FO-SC-45
61	68.29	Very Good	كلية الدراسات الزراعية – البســاتين	AG-38
59	68.29	Fine	كلية الدراسات الزراعية – الاقتصاد الزراعي	AG-29
74	68.29	Good	كلية التربية الفنية	ED-AR-36
66	68.29	Fine	كلية الدراسات الزراعية – البســاتين	AG-38
76	68.29	Good	كلية التربية الفنية	ED-AR-36
60	68.29	Good	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI-40
60	68.29	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
71	68.29	Very Good	كلية التربية الفنية	ED-AR-36
62	68.29	Very Good	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG-52
73	68.29	Good	كلية التربية الفنية	ED-AR-36
60	68.29	Good	كلية التربية الفنية	ED-AR-36
60	68.29	Very Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
82	68.29	Very Good	كلية التربية الفنية	ED-AR-36
69	68.29	Very Good	كلية التربية الفنية	ED-AR-36
59	68.56	Fine	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
65	68.85	Good	كلية علوم وتكنولوجيا الإنتاج الحيواني – الإنتاج الحيواني	ANI-40
57	68.85	Good	كلية الدراسات الزراعية – الهندسة الزراعيـة	AG-80
74	68.85	Fine	كلية الفنون الجميلة والتطبيقية	AR-63

86	68.85	Very Good	كلية التربية الفنية	ED-AR-36
62	68.85	Fine	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
65	68.85	Excellent	كلية الدراسات الزراعية – وقاية النبات	AG-48
59	68.85	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
62	68.85	Very Good	كلية الدراسات الزراعية – الهندسة الزراعية	AG-80
59	68.85	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
61	68.85	Good	كلية علوم الغابات والمراعى – علوم الغابات	FO-SC-45
66	68.85	Excellent	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
63	68.85	Fine	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG-52
77	68.85	Good	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG-52
62	68.85	Fine	كلية الدراسات الزراعية – البساتين	AG-38
60	68.85	Good	كلية الدراسات الزراعية – البساتين	AG-38
66	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
67	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
77	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
69	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
64	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
74	69.56	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
57	69.56	Good	كلية علوم الغابات والمراعى – علوم الغابات	FO-SC-45
60	69.56	Good	كلية علوم الغابات والمراعى – علوم الغابات	FO-SC-45
64	69.56	Fine	كلية الدراسات الزراعية – إنتاج محاصيل حقلية	AG-52
66	69.56	Very Good	كلية الفنون الجميلة والتطبيقية	AR-63
68	69.56	Very Good	كلية الفنون الجميلة والتطبيقية	AR-63
63	69.56	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28
71	69.56	Good	كلية الفنون الجميلة والتطبيقية	AR-63
64	69.56	Good	كلية علوم الغابات والمراعى – علوم الغابات	FO-SC-45
65	69.56	Good	كلية الدراسات الزراعية – إرشاد زراعي وتنمية ريفية	AG-28

74	69.56	Good	كلية التربية البدنية والرياضة	ED-PH-26
63	67.14	Good	كلية علوم الغابات والمراعى - علوم المراعى	FO-SC-39
58	67.14	Very Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
58	67.14	Good	كلية الدراسات الزراعية - إرشاد زراعي وتنمية ريفية	AG-28
64	67.14	Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
62	67.70	Good	كلية علوم الغابات والمراعى - علوم المراعى	FO-SC-39
60	67.70	Good	كلية علوم الغابات والمراعى - علوم الغابات	FO-SC-45
60	67.70	Very Good	كلية الدراسات الزراعية - إرشاد زراعي وتنمية ريفية	AG-28
62	67.70	Very Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
70	67.85	Good	كلية الدراسات الزراعية - الهندسة الزراعية	AG-80
59	67.85	Very Good	كلية علوم المختبرات الطبية - علوم المختبرات الطبية	LA-MED-71
63	67.85	Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
61	67.85	Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
67	67.85	Good	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
65	67.85	Good	كلية الدراسات الزراعية - الهندسة الزراعية	AG-80
62	67.85	Good	كلية علوم الغابات والمراعى - علوم المراعى	FO-SC-39
65	67.85	Good	كلية علوم الغابات والمراعى - علوم الغابات	FO-SC-45
63	67.85	Good	كلية علوم وتكنولوجيا الإنتاج الحيواني - الإنتاج الحيواني	ANI-40
58	67.85	Fine	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
66	68	Good	كلية علوم وتكنولوجيا الإنتاج الحيواني - الإنتاج الحيواني	ANI-40
63	68	Good	كلية الدراسات الزراعية - إرشاد زراعي وتنمية ريفية	AG-28
65	68	Fine	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	AG-52
60	70	Good	كلية علوم وتكنولوجيا الإنتاج الحيواني - الإنتاج الحيواني	ANI-40
66	70	Good	كلية الفنون الجميلة والتطبيقية	AR-63
75	70.14	Good	كلية الموسيقا والدراما - الدراما	MU-27
72	70.14	Excellent	كلية التربية البدنية والرياضة	ED-PH-26
80	70.14	Good	كلية التربية البدنية والرياضة	ED-PH-26

ANI-40	كلية علوم وتكنولوجيا الإنتاج الحيواني - الإنتاج الحيواني	Fine	70.14	64
AR-63	كلية الفنون الجميلة والتطبيقية	Fine	70.14	62
AR-63	كلية الفنون الجميلة والتطبيقية	Good	70.14	68
AG-52	كلية الدراسات الزراعية - إنتاج محاصيل حقلية	Good	70.14	74
AG-29	كلية الدراسات الزراعية - الاقتصاد الزراعي	Excellent	70.14	61
AG-28	كلية الدراسات الزراعية - إرشاد زراعي وتنمية ريفية	Very Good	70.14	61
AG-48	كلية الدراسات الزراعية - وقاية النبات	Good	70.14	65
AG-80	كلية الدراسات الزراعية - الهندسة الزراعية	Good	70.14	64
AR-63	كلية الفنون الجميلة والتطبيقية	Fine	70.14	72

شرح المتغيرات:

- MATH = درجة الطالب في مادة الرياضيات
- Percent = نسبة الطالب في الشهادة السودانية
- RES_DETAIL = تقدير الجامعي للطالب
- FACNAME = اسم الكلية الملنحق بها الطالب
- Namecode =رمز الكلية