١.١ مقدمة

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 وإستغلالها في حل المشاكل واتخاذ القرارت، بإستخدام تطبيقات الحاسوب الحديثة والتي تعتبر تكنولوجيا حديثة ذكية قائمة على جعل الحاسوب "يفكر كما بفكر الانسان ويفعل الانسان" وهو مايعرف بالذكاء الاصطناعي ، جاءت فكرة الكشف والتنقيب علي هذه البيانات بطرق ذكية للمساعدة في حل المشاكل وإتخاذ القرارات.تنقيب البيانات عملية تحليلية تمزج بين علم الذكاء الاصطناعي والاحصاء وتعلم الآلة وقواعد البيانات، وتعتبر خطوة من خطوات إستكشاف المعرفة من قواعد البيانات.

٢.١ مشكلة أهمية البحث

إن أكثر المشكلات والصعوبات التي تواجه تقويم أداء الطالب قد تكون لعدة اسباب منها : ضعف قدرة الوصول إلي تقويم شامل، النتائج المراد تقويمها قد تكون معنوية وتجريدية عامة وليست محسوسة ونوعية خاصة، نتائج التعلم التي نريد قياسها قد يكون من الصعب تحديدها والتعبير عنها بتعابير واضحة ومحددة، لاتزال الكثير من ادوات القياس والتقويم في العلوم الإنسانية غير دقيقة ومبعثاً للشك والإرتياب في صدق ودقة النتائج التي نقيسها مما يؤدي إلى ضعف القدرة في تفسير النتائج وتحليلها .تقييم أداء الطالب يهدف الي توفير القدرة علي الإستكشاف والتركيز علي أهم المعلومات في سجل الطالب، كما يساعد في بناء التنبؤات المستقبلية وإستكشاف السلوك والإتجاهات مما يسمح بتقدير القررات الصحيحة وإتخاذها في الوقت المناسب،كما أنه يمكن يسهم في تعديل معايير الاداء وزيادة أداء الأساتذة والجامعة بشكل عام .

3.1 أهداف البحث

١- إستخدام تقنيات جديده لقياس أداء الطالب.

٢- تحليل البيانات للتوصل لنتائج جديدة.

٣- المساعده في تحسين أداء الطلاب أكاديمياً.

4.1 منهجية البحث

في هذا المشروع سيتم إقتراح طريقة جديدة لتقييم أداء الطالب والمؤسسات التعليمية قائمة علي الطرق المستخدمة حالياً في أنظمة تنقيب البيانات (Data mining systems).هذه الطريقة المقترحة تهدف الي بناء قرارات وتنبؤات ذات صلة بتقييم الطالب مما يقلل من المشاكل الناتجة عن ضعف الطرق المستخدمة في وضع وترتيب المقررات وقد تم إستخدام خوارزمية (aproiri) وتم تطبيقها علي سجل الطالب الأكاديمي.

٥.١ حدود البحث

تشمل الدراسة تطبيق طرق تنقيب البيانات (Data mining) على سجل الطالب، وذلك بإستخدام خوازمية(Aproiri) يتم تطبيقها علي البيانات المأخوذه من إستمارة سجل الطالب وهي عباره عن درجات الطالب المتحصل عليها في مرحله الشهادة الثانوية ،وتشمل الطلاب الذين تم قبولهم سنة (2008–٢٠٠٩) وسنة (٢٠٠٩ – ٢٠١٠) ، وكان النظام التي تم التطبيق علية نظام الساعات المعتمدة وكان الهدف من النتقيب هو تحسين أداء الطالب اكاديميا.

٦.١ هيكلية البحث

يتضمن البحث بالإضافة إلى هذا الباب الابواب:–

الباب الثاني:ويتضمن فصلين نبذة عامة عن تنقيب البيانات ،مهامها ، أنواعها وسجل الطالب

الباب الثالث: ويتضمن الدراسات السابقة.

الباب الرابع: يتناول هذا الباب منهجية البحث.

الباب الخامس:يتناول النتائج ومناقشتها.

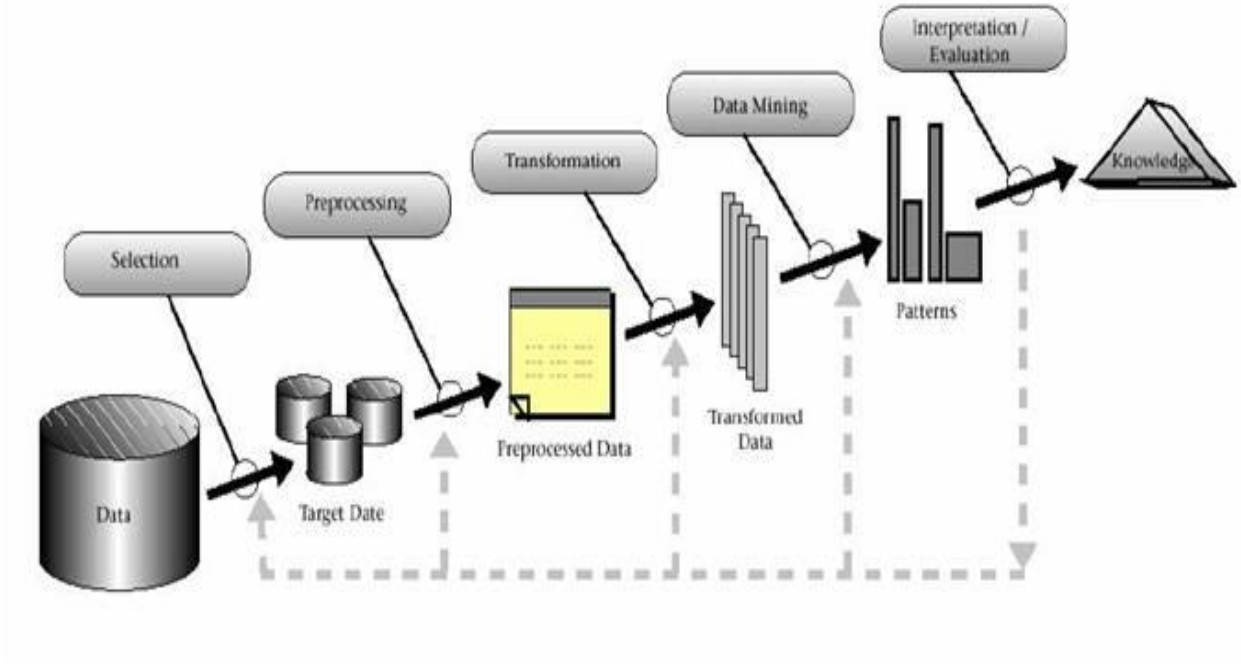
١.٢ تنقيب البيانات

هي عملية بحث عن معرفة من البيانات دون فرضيات مسبقة عما يمكن أن تكون هذه المعرفة. كما ويعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات عادةً ما تكون كمية كبيرة لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطلق إسم (نمودج)على العلاقات والبيانات الملخصة التي يتم الحصول عليها من التنقيب في البيانات . يتعامل تنقيب البيانات عادةً 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 لاتؤثر 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 إستخدام الطرق الإحصائية العادية في تحليلها، و تعتبر عمليه تنقيب البيانات عمليه تحليليه والبحث في بيانات ضخمه وهائله لإستخراج انماط مفيده وإيجاد العلاقات ومدى الإرتباط بين عناصرها.[١]

2.2 أسباب تنقيب البيانات

إن التنقيب في البيانات يهدف الى إكتشاف استخلاص انماط مفيدة،وهي تكنولوجيا مفيدة اصبحت مهمة في ظل التطوير السريع وانتشار إستخدام قواعد البيانات كما أن استخدامها يوفر للمؤسسات واجهزة الامن في جميع المجالات القدرة علي الاستكشاف والتركيز علي أهم المعلومات في قواعد البيانات.تركز تقنيات التنقيب علي بناء تنبؤات مستقبليه واستكشاف السلوك والاتجاهات،مما يسمح بتقدير قرارات صحيحه واتخاذها في الوقت المناسب، غير انها تجيب علي العديد من الاسئله في وقت قياسي خاصة تلك النوعيه من الاسئله التي يصعب الاجابة عليها،ان لم يكن مستحيلا، بأستخدام تقنيات الاحصاء الكلاسيكيه والتي كانت ان وجدت تسترق وقتا طويلا والعديد من اجراءات التحليل.

3.2 خطوات إستكشاف المعرفة من قواعد البيانات(KDD):



شكل(١.٢) يوضح خطوات اكتشاف المعرفة من قواعد البيانات

1.3.2 إختيار البيانات(Data Selection)

في هذه المرحله يتم تحديد واسترجاع البيانات الملائمه من قاعدة البيانات.

2.3.2 تهيئة البيانات(Data Preprocessing)

وهي مرحلة تجهيز وعزل البيانات المبهمه أوالمفقودة أو التي تحتوي على شوائب من بقية البيانات كإلغاء المعلومات المتكررة، التصحيح الشكلي، معالجة البيانات الناقصة وجعلها جاهزة للتطبيق وتشمل هذه المرحله (تنظيف البيانات، إزالة البيانات المفقودة، إشتقاق البيانات، دمج البيانات).

3.3.2 نقل البيانات(Data Transformation)

وهي عملية نقل اليانات التي تم اختيارها الى شكل ملائم لأجراءات البحث والاسترجاع.

4.3.2 تنقيب البيانات(Data Mining)

في هذه المرحله سيتم تطبيق أسلوب ذكي لإستخراج نماذج مفيده بقدر الإمكان.

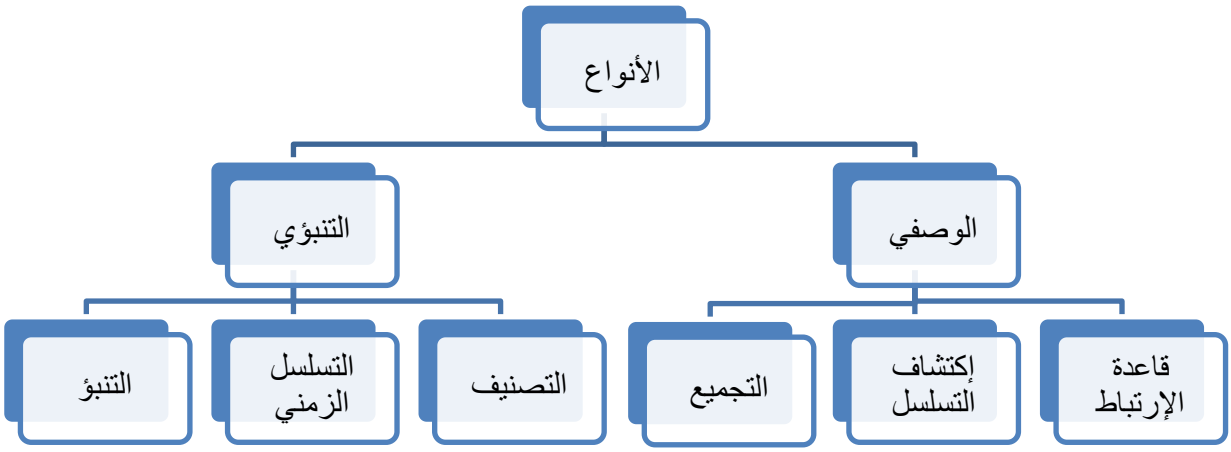
5.3.2 تقييم النموذج(Pattern Evaluation)

بعد إستخراج النماذج المهمه التي تمثل المعرفه يتم تقييمها إستناداً على مقاييس محددة.

6.3.2 تمثيل المعرفة(Knowledge Representation)

وهي المرحله الأخيرة من مراحل إكتشاف المعرفه في قواعد البيانات والتي يراها المستفيد،وهي المرحله الأساسيه والتي تستخدم أسلوب مرئي لمساعدة المستفيد في فهم وتفسير النتائج المستخرجة.

4.2 أنواع تنقيب البيانات



شكل(٢.٢) يوضح أنواع تنقيب البيانات

1.4.2 التنقيب الوصفي

يعتمد علي إعادة تنظيم البيانات، والتنقيب في أعماقها لإستخراج النماذج الموجوده فيها كتشابه الزبائن الذي يسمح لك بإنشاء وصف بسيط عن مجموعه زبائن متشابهين ولايستوجب وجود هدف لمثل تلك التشابه.

2.4.2 التنقيب التنبؤي

يحاول إيجاد أفضل التنبؤات إعتماداً من المعطيات كمعرفه المنتج الأفضل لزبون معين ، بإختصار يعتمد هذا التنقيب على إستخدام معلومات قديمه لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف.

٥.٢ مهام تنقيب البيانات

1.5.2 العنقدة

الهدف منها هو تحديد الاتجاهات داخل البيانات ، ويحاول هذا الاسلوب العثور على مجموعات من العناصر التي توجد عادة معا. وهي عمليه تقسيم البيانات إلى مجموعه من الأصناف إعتماداً على إشتراكها بالخواص المتشابهه وأن العنقدة هي تقسيم غير موجه للبيانات ، وهي عكس التصنيف ، كما أنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

2.5.2 إكتشاف التسلسل

تقنيات التسلسل تستخدم التحليل القائم على الوقت لإنتزاع معلومات مفيدة ، وهي مماثلة للتجميع في انها تستخدم لتحديد العناصر التي تحدث معا ، ولكن الأهم من ذلك انها تستخدم لتحديد أي من العناصر يحدث أولاً.

٣.٥.٢ التصنيف

يستخدم التصنيف بشكل واسع في حل كثير من المشكلات من خلال تحليل مجموعه من البيانات ووضعها في شكل أصناف او أقسام يمكن إستخدامها فيما بعد لتصنيف البيانات المستقبليه وهناك عدد من الطرق التي يمكن إستخدامها في تصنيف البيانات بإستخدام الخوارزميات مثل الخوارزميه الإحصائيه ، وخوارزميات الشبكات العصبيه ، والعديد من خوارزميات تعلم الالة.

٤.٥.٢ الإنحدار

يستخدم للعثور على دالة تنمذج البيانات بأقل خطأ.

٥.٥.٢ التلخيص

توفر عرضا محكما للبيانات يحتوي علي التصور الكامل وتوليد التقرير.

٦.٥.٢ قاعده الإرتباط

هي أحد التقنيات الواعده من تنقيب البيانات كأداه من أدوات إكتشاف المعرفه ولديها القدره على تحليل كميات هائله من البيانات، وهي تسمح بإلتقاط كل القوانين الممكنه التي تشرح بعض الصفات الموجوده إعتماداً علي وجود الصفات الأخرى، بمعني آخر أنها قواعد إرتباطيه معينه بين مجموعه من البيانات في قواعد البيانات.

٦.٢ مقاييس قواعد الإرتباط

هذا التنقيب يكتشف مجموعة من القواعد التي يتم مشاركتها بين نسبة كبيرة من البيانات.وهي تميل الى إنتاج عدد كبير من القواعد والهدف هو العثور على القواعد المفيدة للمستخدمين ، ويمكن ان تقاس الجدوى بطريقتين : موضوعية ، ذاتية.

القياس الموضوعي يشمل التحليل الاحصائي للبيانات مثل confidence،support.

قاعدة الارتباط عبارة عن تعبير ضمني يوضح إرتباط X و Y بالصورة التالية:

$X \rightarrow Y$ حيث ان كل منها تمثل مجموعة البنو(itemset)

Support 1.6.2

هو عبارة عن نسبة توضح جزء من المعاملات التي تحتوي على كل من X و Y.

الجدول الأول : مثال لحساب Support

TID	Items	Support=occurrence/total support
1	ABC	Total support =5
2	ABD	Support{AB}=2/5=40%
3	BC	Support{BC}=3/5=60%
4	AC	Support{ABC}=1/5=20%
5	BCD	

Confidence 2.6.2

تقوم بقياس عدد مرات ظهور العنصر Y في المعاملات التي تحتوي على العنصر X.

الجدول الثاني يوضح:مثال لحساب Confidence:

TID	Items	Given $X \rightarrow Y$

1	ABC	Confidence=Occurrence{Y}/ Occurrence{x}
2	ABD	Confidence{A→B}=2/3=66%
3	BC	Confidence{B→C}=3/4=75%
4	AC	Confidence{A B→C}=1/2=50%
5	BCD	

حيث TID تمثل transaction id .

توجد هنالك العديد من المصطلحات لابد من التعرف عليها لشرح الخوارزمية:

١- مجموعة البنود:عندنا يقال AK- itemset يعني ان تكون مجموعة مكونة من k عنصر.

٢- مجموعة البنود المكررة :هي التي لديها الحد الادنى من support.

١.٢ سجل الطالب

هوبيان يوضح سيرالطالب 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 التقديرالعام 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.

2.2 أهمية سجل الطالب

حفظ المعلومات وتبويبها وتنظيمها في قوالب تساعد على تيسير العمل، كما أنه يساعد علي الإستفادة من معلومات الطالبفي التقييم والتطوير الأكاديمي، بالإضافة الى أنه يستخدم كوسائل توثيق يتم الرجوع إليها عند الحاجة.ملفات الطالب الأكاديمية التي تحتفظ بها الجامعة تقسم الى فئتين عامتين: معلومات الدليل وسجلات الطالب والتي سيتم تعريفها لاحقا.

3.2 سرية السجلات

تتمثل سرية سجل الطالب في أن المعلومات الأكاديمية التي يحتويها هي بالطبع معلومات غير عامة ولا يمكن لأي أحد الوصول اليها إلا للأشخاص المخول لهم بذلك، كما أن تلك المعلومات يمكن الإفصاح عنها وفقا لتوجيهات المسؤول عنها إذا كانت المعلومات هي ضرورية لحماية صحة وسلامة الطلبة او أشخاص آخرين.

4.2 المعلومات الإستدلالية وسجل الطالب

1.4.2 معلومات السجل

يتم تعريفها كالتالي:إسم الطالب، العنوان المحلي ، عنوان البريد الالكتروني ، رقم الهاتف، مكان الميلاد ، المشاركات في النشاطات المختلفة والرياضة ، الهوايات، حالة التسجيل ، الدرجات والمعدلات ، المعاهد التعليمية السابقة للإلتحاق بالجامعة ، عنوان العمل اذا كان الطالب يعمل لد أي جهة ، والكثير من المعلومات المشابهة التي يمكن ان تشمل تصنيف الطلاب والتواريخ المتوقعة لتخرج هؤلاء الطلاب.هذه البيانات الإستدلالية تكون بيانات سريه لا يمكن لاي أحد من الإطلاع عليها الًا الأشخاص المخول لهم ذلك وأيضا لا يمكن الوصول اليه من قبل أي شخص آخر الإ بتقديم خطاب رسمي لإدراة الكليه للسماح لهم

بذلك ويكون ذلك في الحالات الضرورية والتي تتطلب ذلك مثل إحتياج هذه البيانات للمشاريع كما هو الحال في بحثنا هذا.

2.4.2 سجل الطالب

سجل الطالب يشمل السجلات التعليمية التي تحتفظ بها الجامعة بإستثناء معلومات الدليل ويشمل ايضا السجلات الرسمية والشخصية والأكاديمية وتلك المتعلقة بالمسائل المالية والسجلات الدراسية الجامعية . الجامعة لا تسمح بالوصول الى المعلومات الوارده في هذه السجلات دون الحصول على موافقة خطية من إدراة القسم أو الكليه بذلك في حالات الضرورة والحوجة إليها في مشاريع التخرج أو ما شابه ذلك من أسباب كما ذكرنا سابقا.

3.4.2 المستندات الرسمية من السجلات

سيقوم مكتب المسجل بإرسال مستند من هذه السجلات فقط إذا قام الطالب بإرسال خطاب يطلب الحصول عليها.

النسخة الرسمية من سجل كل طالب تحتوي على المعلومات الآتية :

اسم الطالب، تاريخ الميلاد.، اسم ومكان المدرسة التي تخرج منها في الثانوية والسنة التي امتحن فيها، الدورات المتخذة ، عدد الساعات المكتملة في الدراسة ، الدرجات الواردة ، نظام الدرجات والمعدل التراكمي ، سرد تقييمات عن أداء الطالب الأكاديمي وإستخدامها في الحكم علي مدى تقدمه أكاديميا.، تائج اي التماسات او طعون مقدمة من قبل الطالب، إستمارة القبول والوثائق الداعمة، معلومات التسجيل والموقف المالي.

مقدمة

قام الباحثون بإجراء العديد من الأبحاث في ما يخص ملاحظه أداء الطالب الجامعي منذ نهايات القرن الماضي وقد تم تحليل درجات الطالب إحصائيا، وفي الآونة الأخيرة تم إستخدام تقنيات حديثة في تنقيب البيانات في عديد من الدراسات نسرد لكم هذه الدراسات وما ثم من إستخدام تقنيات عليها :

١.٣ استخدم تقنيه التصنيف لتقييم سجل الطالب

قام الباحثين بدراسة في احدى جامعات الهند [٢]بعمل بحث بإستخدام تقنيه التصنيف وتم تطبيقها علي البيانات التعليمية أخذت من درجات طلاب الماجستير وكان عدد الطلاب ٦٠ 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:

١–الكفاءة في توقع نتيجة الطالب.

٢–القيم الايجابية الصحيحة والقيم الايجابية الخاطئة وقيم ال ROC.

٣–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:

1-One-R Algorithm.

2-Decision Table algorithm .

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

٢.٣ إستخدم تقنيه قوانين الإرتباط

دراسة اولى

البيانات التي تطلبت للدراسة في هذا البحث تم أخذها من جامعة مشهوره في منطقة (Haryana)في الهند [٣]لتَحسين أداءالتعليم بتحليل البيانات وإكتشاف العوامل التي تؤثرعلى النتائج الأكاديمية ودراسة أداء الطلاب لكي تزِيد فرص نجاح الطلاب.*وتم عمل هذا البحث علي مجموعه من الطلاب وكان عددهم٣٠ طالب ، وتمت عليها بعض خطوات تجهيز البيانات (*Data Preprocessing*)* *وهي التنظيف (*Cleaning*)* ، *التحويل (*Transformation) وتم تحليل قوانين الإرتباط بأستخدام أحد ادوات تنقيب البيانات وهي (Tanagra)، وذلك بتطبيق قوانين الإرتباط عن طريق (Apriori Algorithm)* لمقارنة أداء طلاب الدراسات العليا في مجال تطبيقات الحاسوب قبل وبعد التخرج.

بالإضافة لتحليل البيانات لمعرفة مدى تأثيرها على النتائج المتوقعة من نجاح او فشل الطلاب ، وبإستخدام هذه التقنيه تم التوصل الى معرفة مدى إهتمام الطلاب بمواد معينة وتصميم المناهج من حيث طريقة التدريس وتقييمها لكي يتم قياس مدى تأثيرها على الطلاب الذين لم ينجحوا لكي يصل بهم الى الحد المقنع من الاداء وكانت قيمه (confidence=16.667)وكانت قيمة (support=83.333)وكان عدد (rules=2) ، بالأخير تم الكشف عن عوامل مختلفة من حيث إهتمام الطالب ، تصميم المناهج ، تقييم المنهجيات التي تؤثر علي الطلاب وتسببت في اخفاقهم.

دراسة ثانية

قام الباحثين بدراسة في جامعة Córdoba في أسبانيا[٤]وتطبيقها على البيانات التربويه في قسم علوم الحاسوب وتم تجميع هذه البيانات بواسطة نظام (Moodle) ووضعت في الجامعة ، وتم إستخدام خوارزمية في قاعدة التجميع بهدف قياس الأداء وإكتشاف العلاقات بين متغيرات البيانات وقد تم تطبيقها علي مجموعة مكونة من ٢٣٠ طالب علي مدار ٥ كورسات من قسم علوم الحاسوب وهذة البيانات تتضمن مجموعة الدرجات التي تحصل عليها الطالب في كل من(assignment, quizzes and forums)وتطبيق كل من خوارزميات RARM وARM.وكان الهدف من الدراسة هو إختبار الأداء والفائدة من تطبيقRARMعلي مجموعة البيانات التعليمية.

وتم تطبيق تقنية قوانين الإرتباط وإستخدام خوارزمية (Apriori) وتطبيق اربعة انواع منها:

.Apriori-Rare −.Apriori-Inverse −.Apriori-Infrequent −.Apriori-Frequent

وكانت المقارنه بين ARMو RARM كالاتي:

Algorithm	# Freq. Itemsets	# UnFreq. Itemsets	# Rules	Avg Support/ ± Std Deviation	Avg Confidence/ ± Std Deviation
Apriori-Frequent	11562	--	788	0.162±0.090	0.717±0.211
Apriori-Infrequent	--	1067	388	0.058±0.060	0.863±0.226
Apriori-Inverse	--	3491	46	0.056±0.070	0.883±0.120
Apriori-Rare	--	5750	44	0.050±0.080	0.885±0.108

جدول (١.٣) يوضح المقارنة بين مقترحات ARM و RARM.

وكانت النتائج المتحصل عليها هي:

Rule	Antecedent	Consequent	Sup	SupC/SupM	Conf
1	total_time_forum=HIGH	mark=PASS	0.24	--/0.47	0.82
2	n_posts=MEDIUM AND n_read=MEDIUM AND n_quiz_a=MEDIUM	mark=PASS	0.13	--/0.25	0.71
3	course=C110 AND n_assignment=HIGH	mark=PASS	0.14	0.52/0.27	0.89
4	total_time_quiz=LOW	mark=FAIL	0.21	--/0.55	0.78
5	n_assignment=LOW	mark=FAIL	0.23	--/0.60	0.70
6	n_quiz_a=LOW AND course=C218	mark=FAIL	0.18	0.51/0.47	0.83

جدول (٢.٣) يوضح القواعد المستخلصة بإستخدام خوارزمية Apriori-Frequent.

١- من القاعدة الأولى نجد ان كل طالب قضى فترة طويلة في المنتدى (قيمة عالية) نجح في الإمتحان النهائي.

٢- من القاعدة الثانية نجد انة اذا كان عدد الرسائل المرسلة والمستقبلة من/الي المنتدى (قيمة متوسطة)وايضا عدد (Quizes) نجح في الإمتحان النهائي.

٣- من القاعدة الثالثة كل طالب أرسل عدد كبير من الواجبات في (course 110) نجح في الامتحان النهائي وكان عدد الواجبات (قيمة عالية) نجح في الإمتحان النهائي.

٤- من القاعدتين الرابعة والخامسة نجد أنه إذا كان مجموع زمن (Quizes) منخفض و عدد (assignment) التي حقق فيها النجاح منخفض ؛ كانت النتيجة رسب في الامتحان النهائي.

٥- من القاعدة السادسة نجد أنه اذا كان عدد (Quizes) في (course 218) منخفض كانت النتيجة رسب في الإمتحان النهائي.

Rule	Antecedent	Consequent	Sup	SupC/SupM	Conf
1	n_quiz=HIGH AND n_quiz_a=HIGH	mark=EXCELLENT	0.045	--/0.69	0.86
2	total_time_assignment=HIGH	mark=EXCELLENT	0.045	--/0.69	0.86
3	n_posts=HIGH AND course=C46	mark=EXCELLENT	0.045	1.00/0.69	1.00
4	total_time_assignment=ZERO AND total_time_forum=ZERO AND total_time_quiz=ZERO]	mark=ABSENT	0.050	--/0.76	0.78
5	n_posts=ZERO AND n_read=ZERO	mark=ABSENT	0.050	--/0.76	0.78
6	n_quiz=ZERO AND course=C111	mark=ABSENT	0.050	0.88/0.76	1.00

جدول (٣.٣) يوضح القواعد المستخلصة بإستخدام خوارزمية Apriori-Rare

١- من القاعدة الأولى نجد ان اذا كان عدد (Quizzes)كبير وعدد (Quizzes) التي نجح فيها الطالب كبير يحرز درجة الإمتياز في الإمتحان النهائي.

٢- من القاعدة الثانية نجد ان الطالب الذي يقضي وقت طويل في الواجبات يحرز درجة الإمتياز.

٣- من القاعدة الثالثة نجد ان الطالب الذي يرسل مجموعة كبيرة من الرسائل في (course 46) إلي المنتدى يحرز درجة الامتياز.

٤- القواعد الثلاث الأخيرة خاصة بالطلاب الغائبين عن الإمتحان النهائي ، حيث أنه اي طالب لم تكن له درجات في الواجبات ، الإختبارات ومشاركات المنتدى لم يجلس للإمتحان النهائي.

٣.٣ استخدم تقنيه التنبؤ

في هذا البحث نقوم بالتنبؤ المبكر (Early Prediction) من نتيجة الدراسة [٥]التي تستعمل بيانات تسجيل الطالب ، والهدف من ذلك البحث هو إكتشاف العوامل التي قد تؤثر على نتائج الدراسة في نظم المعلومات في كلية الفنون التطبيقية المفتوحة في منطقة نيوزلندا ، وكان عدد البيانات المستخدمة ٤٥٣ طالب.

تم إستخدام بيانات القيد (Enrolment Data) لتحقيق الاهداف الأتية:

١- بناء نماذج للتنبؤ المبكر من نتائج الدراسة بإستخدام بيانات القيد.

٢- تقييم النماذج بإستخدام (Cross-Validation)و(Misclassification) وذلك لإتخاذ القرار.

٣- النتائج الحالية يسهل فهمها من قبل المستخدمين.

كانت بيانات القيد تخص المعلومات الآتيه:

الجنس ، تاريخ الميلاد، الأصول العرقية والعجز، حالة العمل ، تأهيل المدارس الثانويه الأكاديمية، Programmingcourse، 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:

١- CHAID .

٢- Exhaustive CHAID.

٣- QUEST.

٤- CART.

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

١- الاصول العرقية.

٢- Programming Course.

٣- Course block.

ولكن لسوء الحظ أشجار التصنيف لم تعطى دقة عالية في النتائج ، وكان دقة شجرةCHAID هي ٥٩.٤% وكانت دقة شجرة CART هي 60.5% ، وهي تقترح أن المعلومات المساعدةَ (الجنس،العُمر ، الإنتماءالعرقي ، 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.

مقدمة

بالتعامل مع سجل الطالب الأكاديمي نجد أنه تتوفر الكثير من المعلومات الأكاديمة المتعلقة بالطالب وهي غير مستغلة ، وفي هذا البحث تم تحديد بعض هذه البيانات والتعامل معها وقد تمت معالجتها وفقا لخطوات تنقيب البيانات وهي كالتالي :

1.4 أولاًإختيار البيانات

تم الحصول على البيانات من قبل مكتب المسجل في الكلية وهي تخص قسم علوم الحاسوب ، وتتمثل هذه البيانات في نتائج القبول من المرحلة الثانوية و بيانات نتائج الطالب التي تحصل عليها خلال دراسته الجامعية على مدار الأربع سنوات . وقد تحصلنا في بداية تجميع البيانات علي٩٧ طالب وطالبة من قسم علوم الحاسوب ولكن بعد المعالجات التي تمت عليها قل هذا العدد الى ان تم التعامل مع ٣٤ طالب وطالبة.

وعند إدخال البيانات ظهرت بعض المشاكل التي تخص الطلاب أصحاب الشهادة العربية لأن الطلاب والطالبات القادمين من الدول العربية الأخرى يكونون قد جلسو لمواد لا تدرس هنا في المقررات السودانية ،لذلك تعتبر شهادتهم مختلفة عن شهادتنا ولا تحتوي على نفس التفاصيل ، وطلاب التجسير حيث أنه لا تتوفر لهم بيانات الشهادة الثانوية السودانية.

2.4 ثانيا تهيئة البيانات

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. وكانت مكتوبه في شكل(Excel sheet) وكان الملف يحتوي على جميع المواد خلال الأربع سنين ، وكل ماده تحتوي على جميع التفاصيل للماده العملي والنظري وأعمال السنة والنتيجة النهائيه والتقدير، حيث تم فصل الطلاب الذين تعذر الحصول علي درجاتهم الخاصة بنتائج الشهادة الثانوية وتخزينهم منفصلين عن بقية الطلاب ، بالإضافة الي تهيئة الشكل العام للملف بحيث يناسب أداة تنقيب البيانات المستخدمة وهي (WEKA)وهي تقبل البيانات في صورة ملف (ARFF) أو قاعدة بيانات.

تم نقل الدرجات الخاصة بطلاب علوم الحاسوب الدفعة(٢٠٠٨-٢٠٠٧) للشهادة السودانية من إستمارات التسجيل المخزنة في مكتب المسجل ، وتحويلها الى(Excel sheet)بالإضافة الي البيانات التي تم أخذها من قسم علوم الحاسوب والتي تتضمن كافة الدرجات التفصيلية الخاصة بالمواد علي مدار الأربع سنين ودرجات مشروع التخرج وتحويلها الي (Excel sheet) أيضا كما تم التحصل علي المعلومات الشخصية للطالب مثل(النوع ، العنوان وتاريخ الميلاد) من الشؤون العلمية ودمجها مع البيانات المشار اليها مسبقاً.

3.4 ثالثا تنقيب البيانات

في هذه المرحلة تم إستخدام تقنية قاعدة الإرتباط وتطبيق خوارزمية (Aproiri) بإستخدام برنامج (Weka).

4.4 ما تحتويه Weka

يمدنا برنامج (Weka) [١]بمجموعة من خوارزميات تعلم الآلة والتي تستخدم لإجراء مهام تنقيب البيانات.هذه الخوارزميات يمكن أن تطبق مباشرة على مجموعة بيانات أو يتم استدعاؤها من التعليمات البرمجية الخاصة بالجافا ، ويحتوي أيضا على أدوات تجهيز البيانات بالطرق المختلفة وكذلك أدوات تحويل قواعد البيانات مثل الخوارزميات .

5.4 مقدمة عن weka

طاولة العمل تحتوي على الأدوات المستخدمة في المشاكل الرئيسية في تنقيب البيانات مثل :الإنحدار والتصنيف والتجميع ، جميع الخوارزميات تأخذ مدخلاتها في شكل جدول أحادي العلاقة في صورة (ARFF).

6.4 صيغة ARFF

الصفوف التي تبدأ بعلامة (%)هي تعليقات ، يلي التعليقات في بداية الملف هو إسم العلاقة ، الخواص الإسمية يجب ان يتبعها مجموعة من القيم التي يمكن ان تأخذها وتكون هذه القيم متضمنة في الأقواس المعرجة. القيم يمكن أن تحتوي علي فراغات في هذه الحالة لابد من ان تكون مضمنة بين علامات تقدير،الخواص الرقمية لابد ان تليها الكلمة المحجوزة (Numeric)كما هو موضح بالشكل التالي الذي يوضح مثال لملف (ARFF) للعلاقة (project):

```
@relation project
@attribute sn numeric
```

@attribute sex numeric
@attribute Arc numeric
@attribute ENC numeric
@attribute physics numeric
@attribute chemistry numeric
@attribute biology numeric
@attribute compute numeric
@attribute engineering numeric

أحد الطرق لإستخدام (Weka) هي تطبيق طريقة التعلم لمجموعة البيانات وتحليل مخرجاتها لمعرفة المزيد عن البيانات، وايضاً يمكن إستخدام النماذج التي تم تعليمها مسبقا لتوليد التوقعات على حالات جديدة، وكذلك يمكن إستخدام مجموعة مختلفة من النماذج التي تم تعليمها مسبقاً ومقارنة اداؤها من أجل إختيار احدها للتنبؤ.

في واجهة (Weka) التفاعلية يمكن تحديد طريقة التعلم المراد إستخدامها من القائمة ، والكثير من الطرق لديها مجموعة من المعلمات الإنضباطية التي يمكنك الوصول اليها من خلال ورقة الملكية أو من خلال محرر كائن، ويتم إستخدام وحدة التقييم المشتركة لقياس أداء جميع المصنفات.تطبيقات نظم التعلم الفعلي هي من المصادر الأكثر قيمة التي توفرها (weka)،ولكن الأدوات اللازمة لتجهيز البيانات التي تسمى المرشحات تأتي في المرتبة الثانية.مثل المصنفات يمكنك تحديد المرشحات من القائمة وتكييفها مع متطلباتك.

7.4 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 (ARFF) 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 للـ(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.

8.4 البداية

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.

9.4 تهيئة البيانات(Weka)

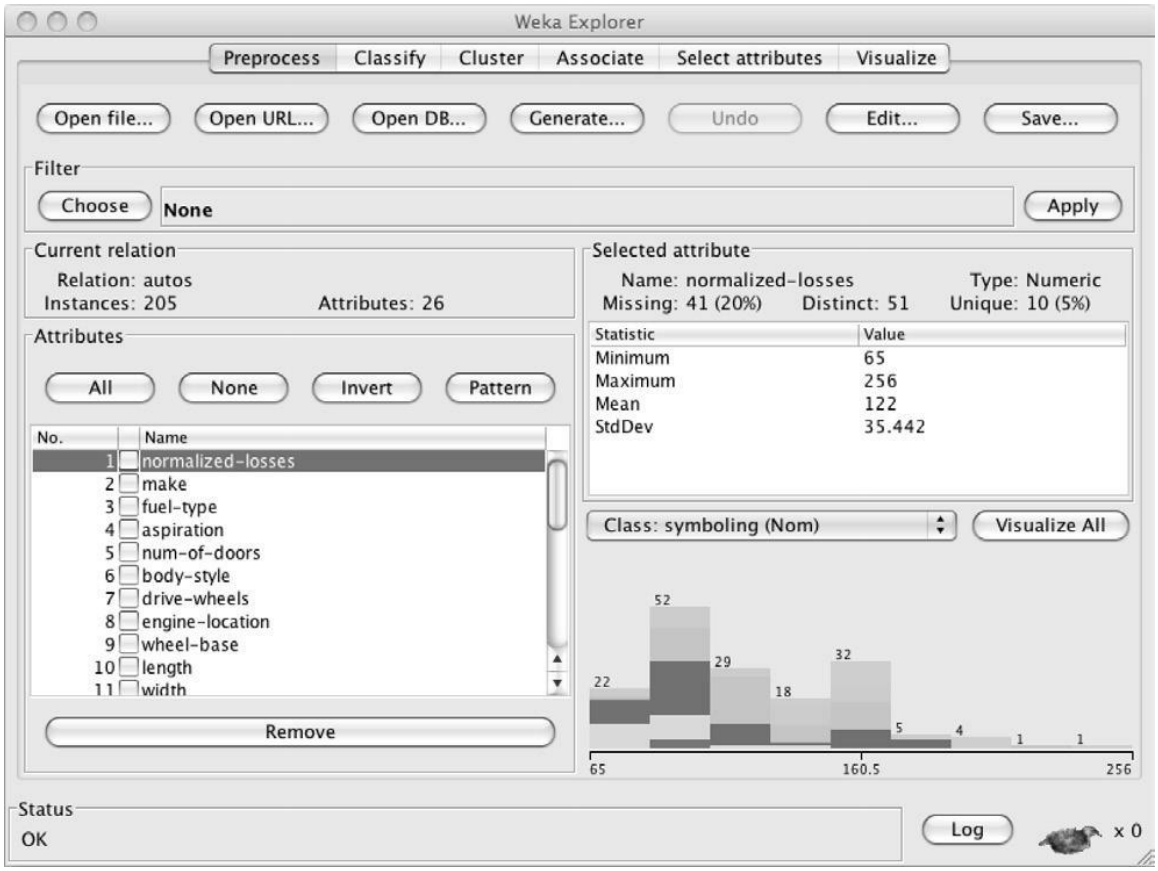
غالبا ما يتم عرض البيانات في صورة (Excel Sheet)أو قاعدة بيانات ، على أي حال الطريقة الأصلية للتخزين في (Weka)هي في صورة (ARFF)ويمكن بكل سهولة التحويل من صورة برنامج الجدولة الى صورة (ARFF).معظم ملفات (ARFF)تتكون من قائمة من الحالات وقيم خواص لكل من الحالات التي تم الفصل بينها بالفاصلة كما ذكرنا سابقا.معظم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

10.4 تحميل البيانات في المستكشف

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

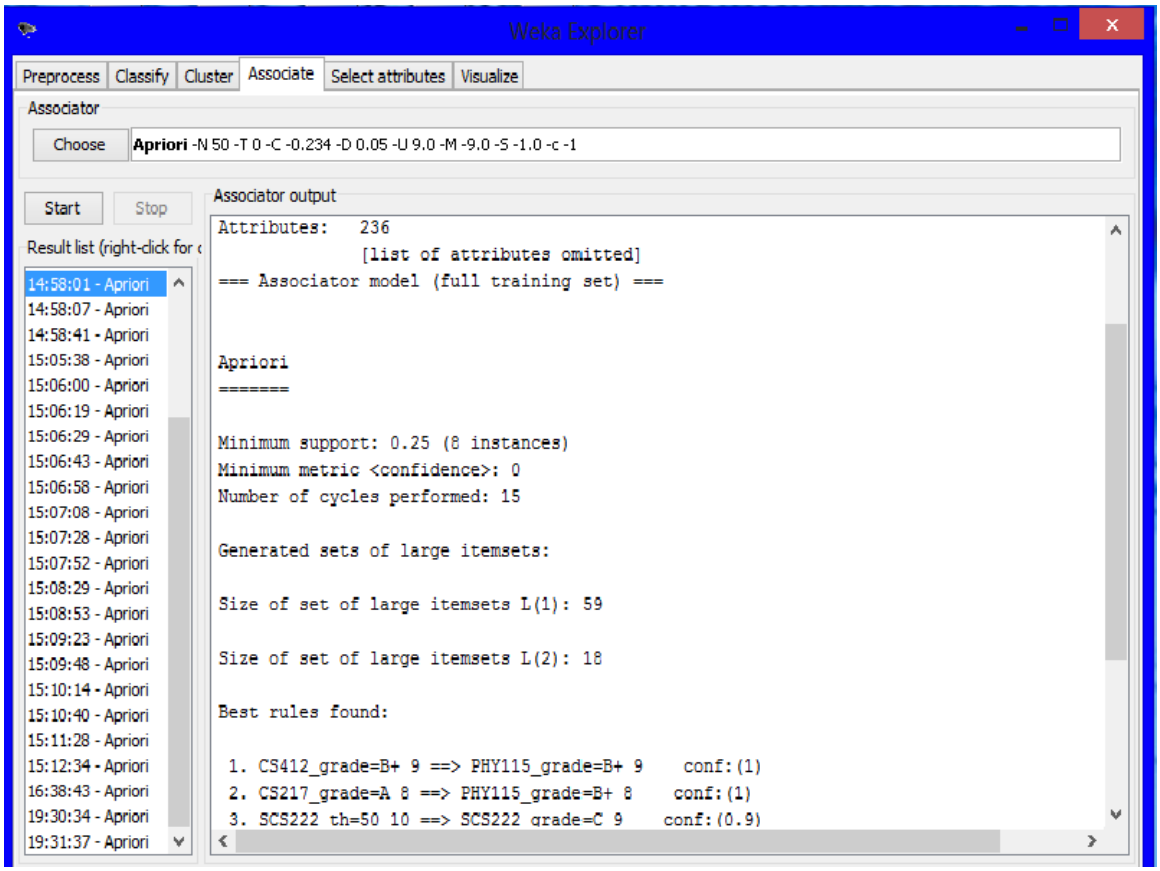
الشكل (1.4) : واجهة الاختيار الرسومية.

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



الشكل (٢.٤) : 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 (ARFF).بعد ان يتم تحميل البيانات نقوم بتحديد الصفات(attributes) التي نريد إجراء عملية التجميع عليها ويوجد أيضا خيار إختيارها جميعها وإجراء التجميع عليها. ومن ثم بالرجوع للست قوائم الأولى نقوم بإختيار associate ثم start ، بعدها يقوم البرنامج بعمل توليد للنتائج والمخرجات كما هو موضح في الشكل (3.4) :



الشكل (٣.٤) : شكل الشاشة عندما يتم الحصول على النتائج.

٤.١١ خوارزمية Aproiri

هي خوارزمية للتنقيب عن العناصر المتكررة وتطبيق قوانين الإرتباطعلى(transactional databases).وتعمل عن طريق تحديد العناصر الفردية المتكررة في قاعده البيانات وتوسيعها لتشمل مجموعات أكبر. وأكبر بند هوالبند الأكثر ظهوراً في قاعدة البيانات ، ومجموعات البند المتكررة (The frequent item sets) يتم تحديدها عن طريق خوارزمية (Aproiri) التي تستخدم لتحديد قوانين

الإرتباط.هذه الخوارزمية مصممة للعمل على قواعد البيانات التي تحتوي على (transactions) ، وتم تصميم الخوارزميات الأخرى لإيجاد قوانين الإرتباط في حالة عدم وجود(transaction)أو عدم وجود (timestamps) وينظر إلى كل (transaction) علي أنها مجموعة من الوحدات(item set) ، تستخدم الخوارزمية منهج "bottom up" حيث يتم تمديد مجموعات فرعيه للعنصر الواحد في اللحظه المعينة وهي خطوة(candidate generation) ويتم إختبار هذه المجموعات المرشحة وتنتهي هذة الخوارزمية عندما يتم العثور على أي تمديدات أخرى ناجحة.وتستخدم هذة الخوارزمية كل من(breadth-first search)و(Hash tree) لحساب المجموعات المرشحة بكفاءة ، حيث يتم توليد المجموعات المرشحة ذات الطول k من المجموعات المرشحة ذات الطول k-1.

الخوارزمية كالآتي:

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\qquad \textbf{while } L_{k-1} \neq \emptyset\\
&\qquad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}\\
&\qquad\quad \textbf{for transactions } t \in T\\
&\qquad\qquad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\qquad\qquad \textbf{for candidates } c \in C_t\\
&\qquad\qquad\quad count[c] \leftarrow count[c] + 1\\
&\qquad\quad L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}\\
&\qquad\quad k \leftarrow k + 1\\
&\qquad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

توضيح الرموز:

١. (T) توضح Transaction.

٢. (€) توضح support.

٣. (Cк)توضح المجموعة المرشحة في المستوىk .

٤. (Count[c])توضح حقل هيكلة البيانات التي تمثل المجموعات المرشحة c ،الذي من المفترض أن تكون قيمتة الإبتدائية تساوي صفرا.

٥. (L) توضح اكبر بند (Large item set).

٥.١مقدمة

بعد إجراء العديد من التجارب علي برنامج (Weka) وذلك بتغيير قيم (Confidence)و (Support) ،وتغيير الصفات التي يتم إختيارها في مرحلة التهيئة للبيانات (Preprocessing) تم التحصل على كمية كبيرة منالقواعد (Rules).وبعد تحليل تلك القواعد وجد أن

بعضها يصنف على أنه بديهي(Trivial) والبعض الأخر تتباين درجة منطقيته علي أساس قيمة ال(Confidence).

بعد تحليل النتائج المتحصل عليها كافة قمنا بإختيار القواعد التى توضح قوة إرتباط المواد ببعضها من حيث التأثير علي مستوي الطالب بصورة واضحة مما يساعد في إتخاذ القرارات الملائمة لتطوير وتحسين المستويات الأكاديمية للطلاب عن طريق إتباع منهجية معينة (كزيادة عدد ساعات مادة معينة أو العكس).

الجدول (١.٥) يوضح ملخص النتائج المتحصل عليها بعد تطبيق خوارزمية (Aproiri) موضحا عدد المجموعات المتكررة(Frequent item sets)التي تم تنقيبها وعدد القواعد المكتشفة ومتوسط كل من (confidence and support) لكل قاعدة :

n.item sets	Large item set	n.Rules	Support	Confidence
٢	٥	10	0.1	0.9
1	7	100	0.01	0.2
1	6	200	٠.٠٠١٠	0.2

جدول (١.٥) : ملخص النتائج المتحصل عليها

ومن النتائج لاحظنا انه كلما زادت نسبة (support)يقل عدد (item set) وكلما كانت عدد (item set) الناتجة قليل كلما كانت العلاقات بينهم منطقية اكثر،التغيير في كلاً من قيم (confidence) و (support) بزيادة أو نقصان يؤدي إلى ظهور قواعد مختلفة ذات درجات مختلفة من المنطقية.

كما ظهرت كمية كبيرة من النتائج البديهية والنتائج الغير منطقية عند تحديد جميع الصفات لإيجاد العلاقات بينهم مما جعلنا نقوم بتحديد صفات معينة واجراء العمليات عليها لاكتشاف علاقات اكثر منطقية بين المواد.

```
19. PHY115_grade=B+ CS317_grade=B+ 8 ==> CS217_grade=A 7    conf:(0.88)
20. PHY115_grade=B+ CS217_grade=A 8 ==> CS317_grade=B+ 7    conf:(0.88)
21. CS217_grade=A 8 ==> PHY115_grade=B+ CS317_grade=B+ 7    conf:(0.88)
22. PHY115_grade=B+ CS325_grade=B+ 8 ==> CS217_grade=A 7    conf:(0.88)
23. PHY115_grade=B+ CS217_grade=A 8 ==> CS325_grade=B+ 7    conf:(0.88)
24. CS217_grade=A 8 ==> PHY115_grade=B+ CS325_grade=B+ 7    conf:(0.88)
25. PHY115_grade=B+ CS317_grade=B+ 8 ==> CS412_grade=B+ 7    conf:(0.88)
26. PHY115_grade=B+ CS325_grade=B+ 8 ==> CS412_grade=B+ 7    conf:(0.88)
27. PHY125_grade=B 12 ==> sex=1 10    conf:(0.83)
28. CS323_grade=B 10 ==> sex=1 8    conf:(0.8)
29. ARB121_grade=A 10 ==> ISC113_grade=B+ 8    conf:(0.8)
30. CS325_grade=B+ 10 ==> PHY115_grade=B+ 8    conf:(0.8)
31. CS325_grade=B+ 10 ==> CS314_Grade=B+ 8    conf:(0.8)
32. ARB111_grade=C 9 ==> sex=1 7    conf:(0.78)
33. CS117_grade=A 9 ==> sex=1 7    conf:(0.78)
34. PRG415_grade=A+ 9 ==> sex=1 7    conf:(0.78)
35. CS317_grade=B+ 9 ==> sex=1 7    conf:(0.78)
36. EN322_grade=C 9 ==> ISC113_grade=B+ 7    conf:(0.78)
37. SCS311_grade=B+ 9 ==> ISC123_grade=B+ 7    conf:(0.78)
38. CS412_grade=B+ 9 ==> CS217_grade=A 7    conf:(0.78)
39. CS317_grade=B+ 9 ==> CS217_grade=A 7    conf:(0.78)
40. EN322_grade=C 9 ==> PRG415_grade=A 7    conf:(0.78)
41. CS317_grade=B+ 9 ==> CS412_grade=B+ 7    conf:(0.78)
42. CS412_grade=B+ 9 ==> CS317_grade=B+ 7    conf:(0.78)
43. CS412_grade=B+ 9 ==> CS325_grade=B+ 7    conf:(0.78)
44. CS317_grade=C 9 ==> CS324_grade=C 7    conf:(0.78)
45. CS324_grade=C 9 ==> CS317_grade=C 7    conf:(0.78)
46. CS317_grade=B+ 9 ==> CS325_grade=B+ 7    conf:(0.78)
47. sex=2 ISC113_grade=B+ 9 ==> PRG415_grade=A 7    conf:(0.78)
48. PHY115_grade=B+ CS412_grade=B+ 9 ==> CS217_grade=A 7    conf:(0.78)
49. CS412_grade=B+ 9 ==> PHY115_grade=B+ CS217_grade=A 7    conf:(0.78)
50. CS317_grade=B+ 9 ==> PHY115_grade=B+ CS217_grade=A 7    conf:(0.78)
```

شكل (١.٥):جزء من النتائج المتحصل عليها.

وكانت القواعد المعتمدة كالأتي:

1. CS216_grade=A+ 5 ==> MAT114_grade=A
2. CS315_grade=A 5 ==> CS217_grade=A
3. CS226_grade=A 5 ==> CS315_lab=50 5
4. CS117_grade=A CS217_grade=A 5 ==> CS317_grade=B+ 5
5. CS422_grade=D 6 ==> CS227_grade=C 5
6. CS227_assignment=50 6 ==> CS315_lab=50 5
7. CS227_grade=A 6 ==> CS217_grade=A 5
8. CS422_grade=D 6 ==> CS127_grade=D 5
9. Mathematics=88 3 ==> MAT114_grade=A 3
10. Mathematics=88 3 ==> PHY115_assingment=37 3
11. MAT214_grade=D 6 ==> MAT124_grade=D 5

٢.٥ مناقشة النتائج:

١- الطلاب الذين يحرزون التقدير ممتاز(+A) في مادة الهياكل المتقطعة I يحرزون التقدير ممتاز(A) في مادة الرياضيات I.

٢- الطلاب الذين يحرزون التقدير ممتاز (A) في مادة الرسم بالحاسوب كانو قد أحرزوا التقديرممتاز(A) في مادة أساليب البرمجة II.

٣- الطلاب الذين يحرزون التقدير ممتاز(A) في مادة هياكل البيانات والخوارزميات كانو قد أحرزوا الدرجة ٥٠ في معمل مادة الرسم بالحاسوب.

٤- الطلاب الذين يحرزون التقدير ممتاز(A) في مادة مدخل لعلوم الحاسوب ونفس التقدير في مادة أساليب البرمجة I يحرزون التقدير جيد جدًا(B) في مادة الذكاء الاصطناعي.

٥- الطلاب الذين يحرزون التقدير مقبول (D) في مادة مفاهيم لغات البرمجة كانو قد أحرزوا التقدير جيد (C) في مادة اساليب البرمجة II.

٦- الطلاب الذين أحرزوا التقديرممتاز (A) وكانت درجة واجباتهم ٥٠ في مادة أساليب البرمجةII نجحوا في معمل مادة الرسم بالحاسوب.

٧- الطلاب الذين أحرزوا التقدير ممتاز(A) في مادة اساليب البرمجة I يحرزوا نفس التقدير في اساليب البرمجةII.

٨- الطلاب الذين يحرزوا التقدير مقبول(D) في مادة مدخل لعلوم الحاسوب يحرزون نفس التقدير في مادة مفاهيم لغات البرمجة.

٩- الطلاب الذين يحرزون الدرجه ٨٨ في مادة الرياضات في المرحلة الثانوية يحرزون التقدير ممتاز(A) في مادة الرياضيات I.

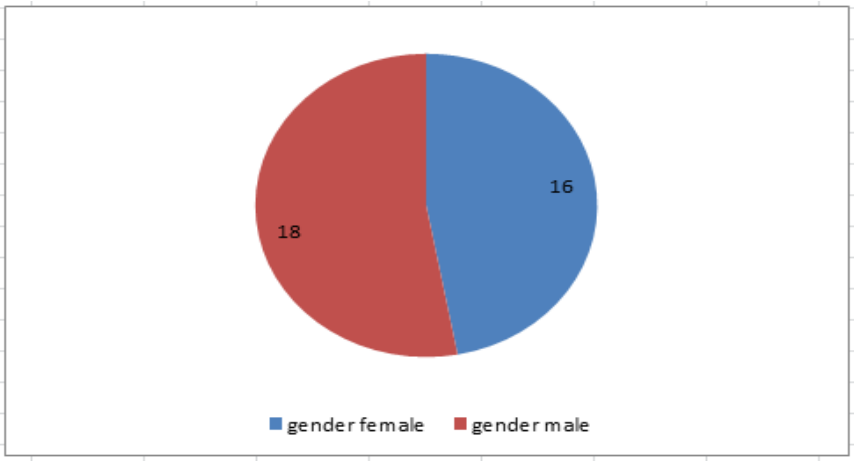
١٠-الطلاب الذين أحرزوا الدرجة ٨٨ في مادة الرياضيات في المرحلة الثانوية ينجحون في واجبات مادة الفيزياءI.

١١-الطلاب الذين أحرزوا التقدير مقبول (D) في مادة الرياضيات III كانو قد أحرزوا نفس التقدير في مادة الرياضيات II.

٣.٥ إحصائيات البيانات

قمنا بعمل مجموعة من المخططات بناءً علي البيانات المتحصل عليها وهي توضح الأداء في كل المواد علي مدار الأربع سنين،بالإضافة إلى نتيجة الشهادة الثانوية وتفاصيلها كما هو موضح ادناه:

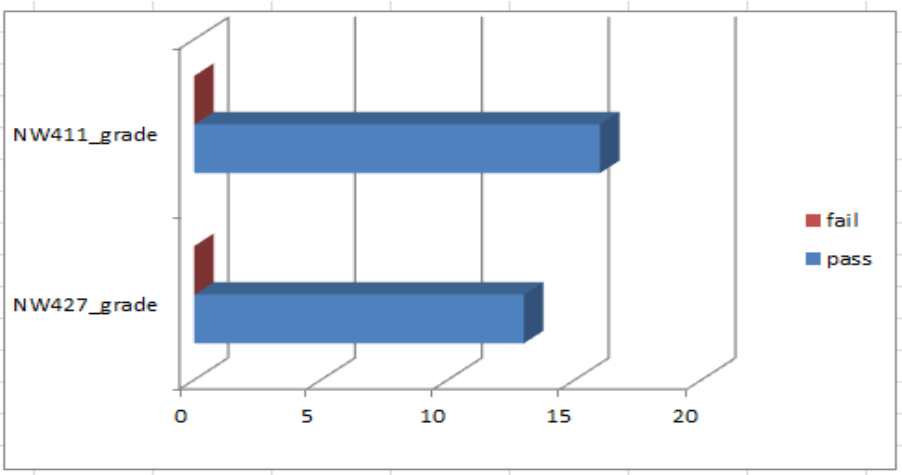
1.3.5مخطط يوضح نسبة الذكور الى الاناث



شكل(٢.٥):مخطط يوضح نسبة الذكور للإناث

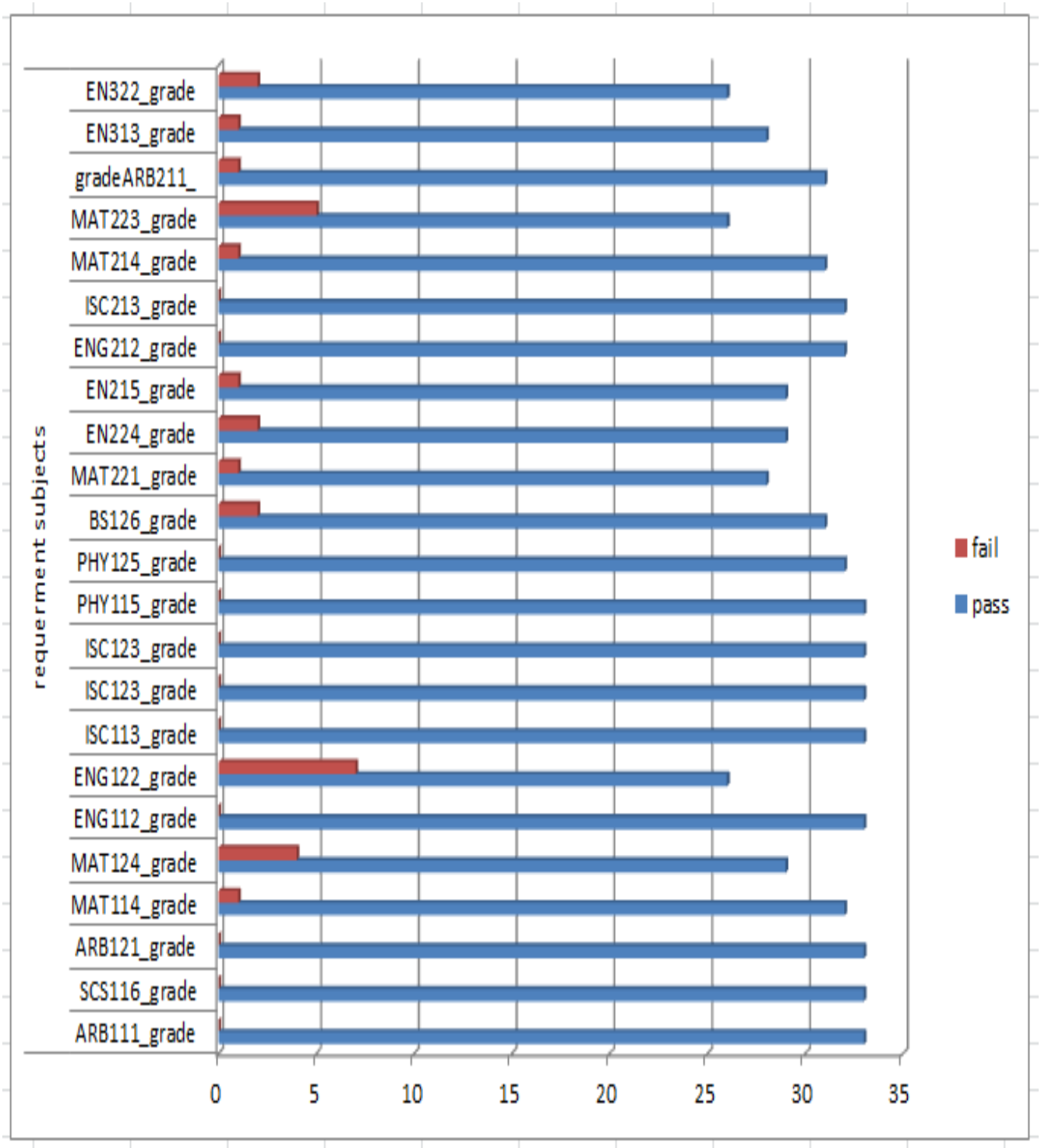
2.3.5 مخططات توضح المواد التي تتبع لنفس التخصص والمتطلبات:

- **قسم الشبكات :**



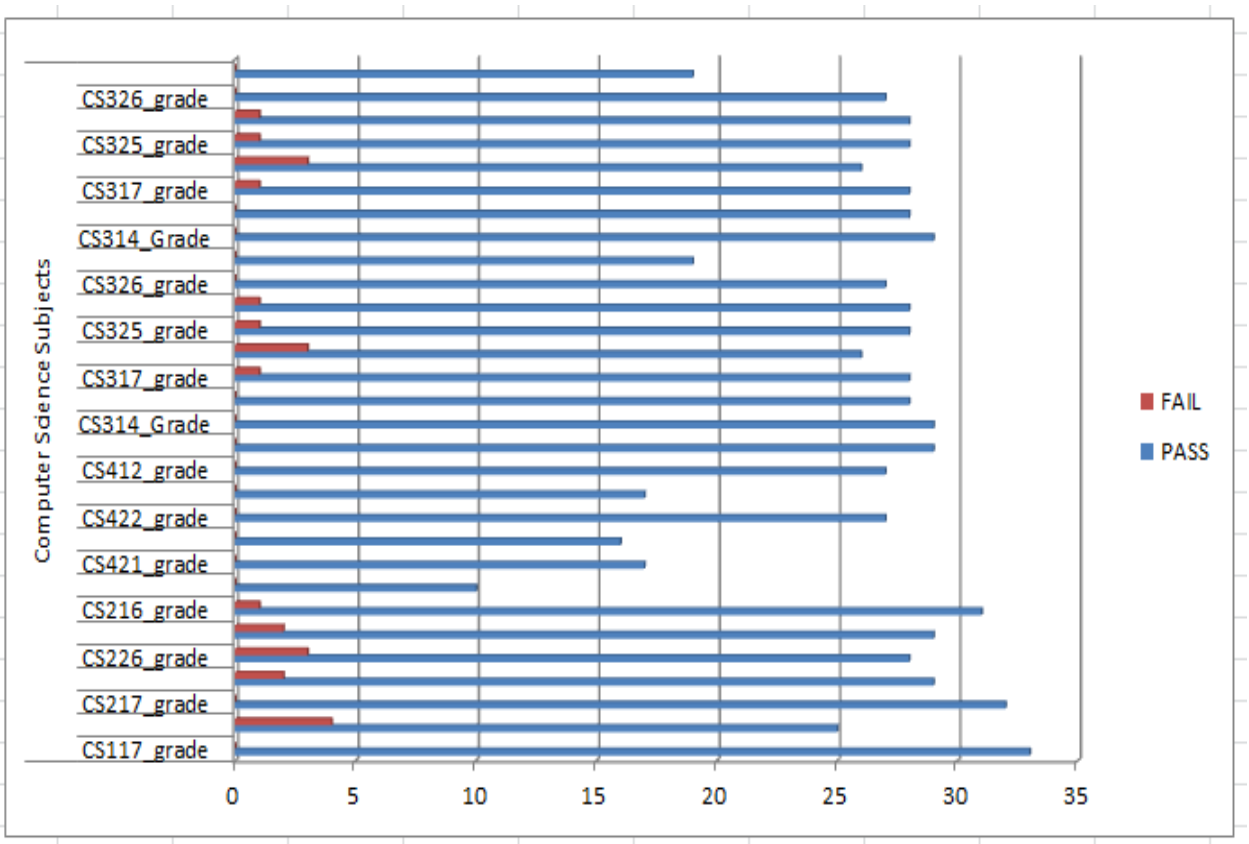
شكل(٣.٥) الأداء في المواد التي تمثل مواد التخصص لقسم الحاسوب والشبكات

- **متطلبات الكلية :**



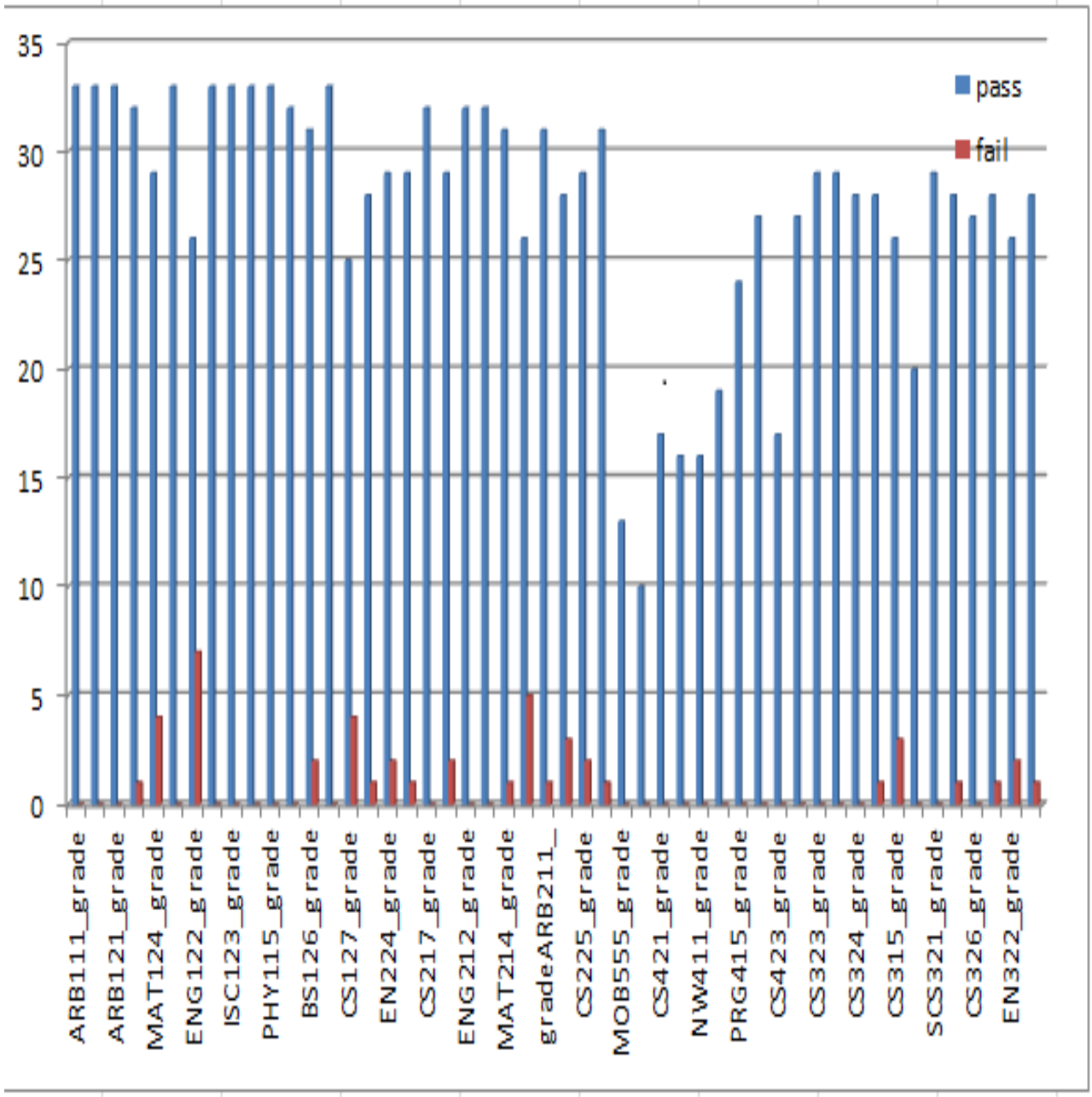
شكل(٤.٥):الأداء في المواد التي تمثل متطلبات الكلية.

- **قسم علوم الحاسوب:**



شكل (٣.٥):الأداء في المواد التي تمثل التخصص في قسم علوم الحاسوب.

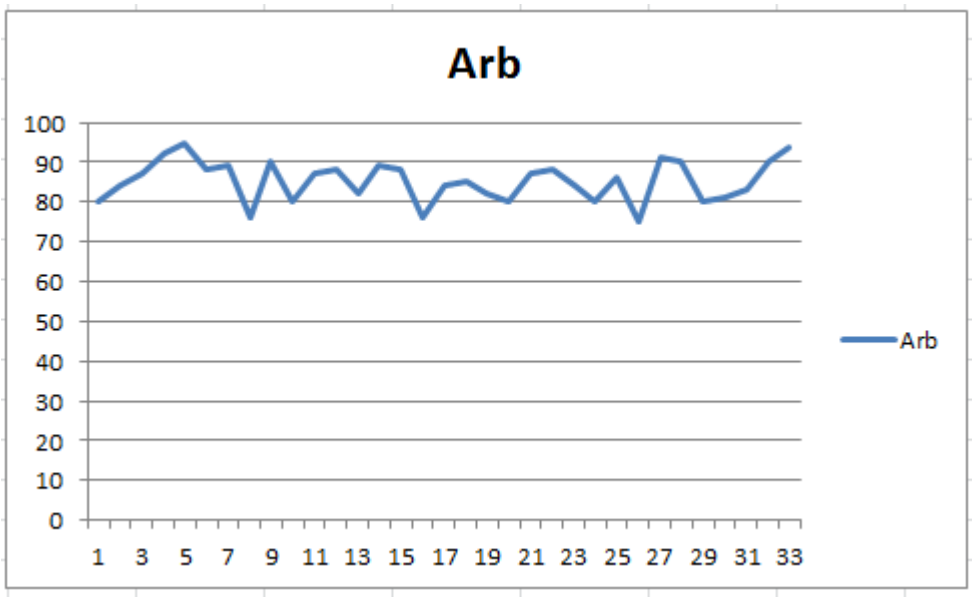
3.3.5 مخطط يوضح العلاقات بين المواد على مدار الأربع سنوات:



شك(٦.٥):العلاقة بين المواد الجامعية.

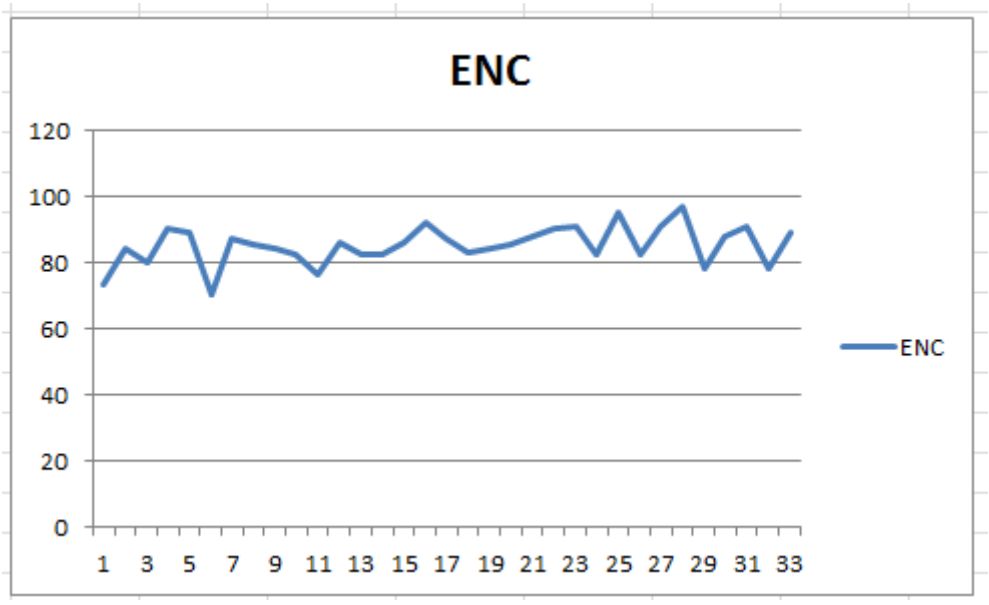
٤.٣.٥ تباين درجات الطلاب في مواد الشهادة الثانوية

مادة اللغة العربية



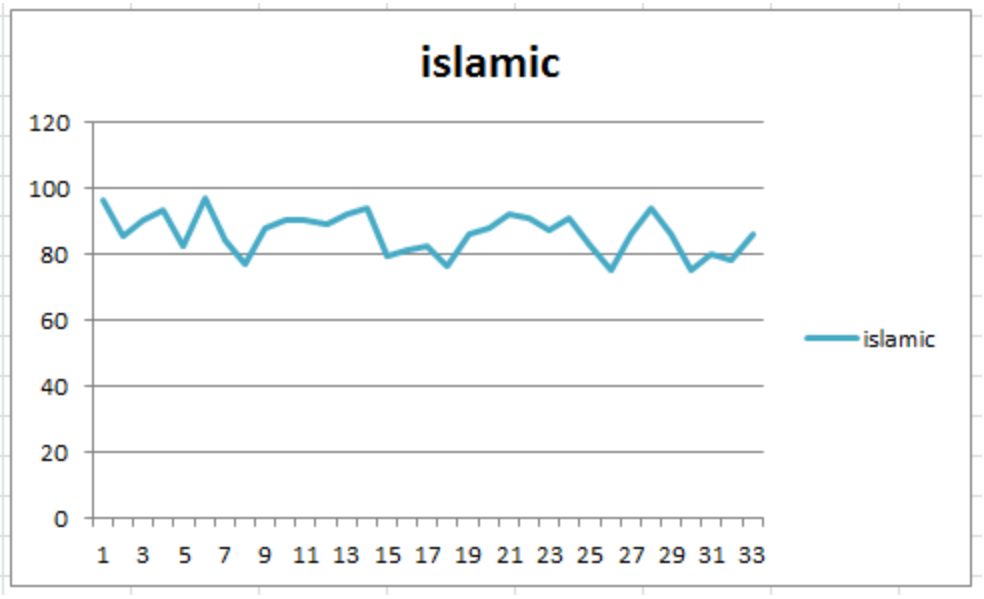
شكل(٥.٧): مادة اللغة العربة.

مادة اللغة الإنجليزية



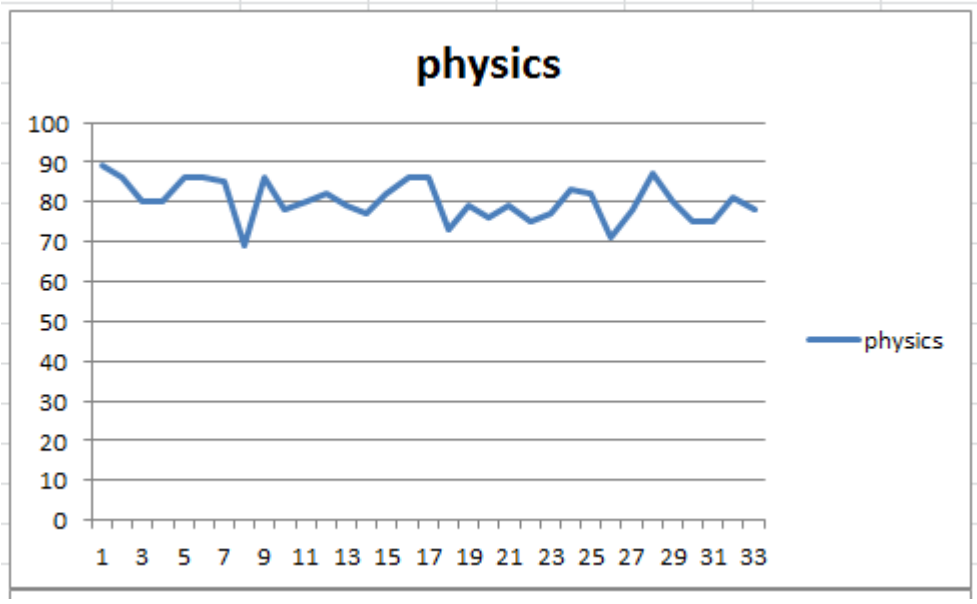
شكل(٥.٨): مادة اللغة الإنجليزية.

مادة الدراسات الإسلامية



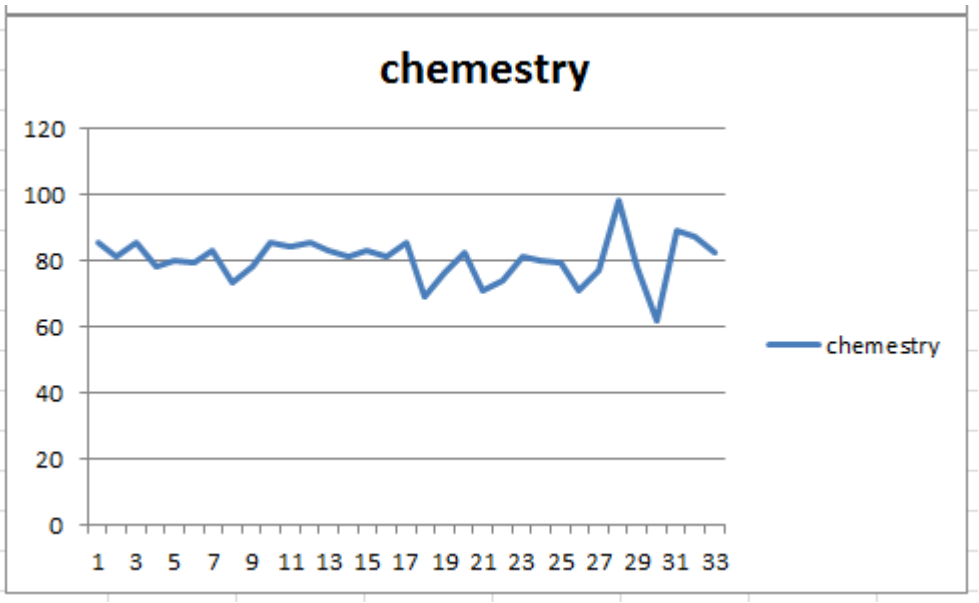
شكل(٥.٩): مادة الدراسات الإسلامية.

مادة الفيزياء



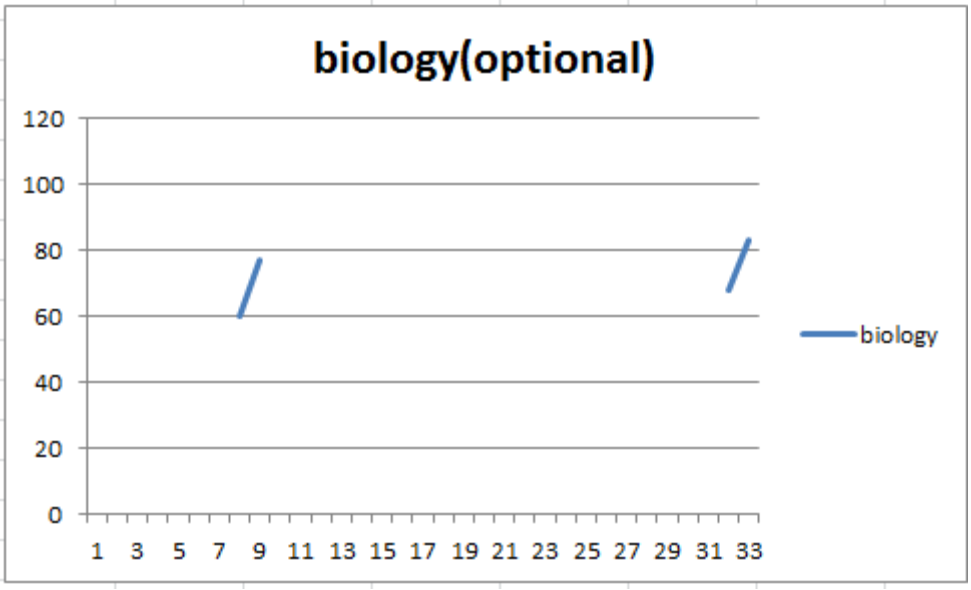
شكل(٥.١٠): مادة الفيزياء.

مادة الكيمياء



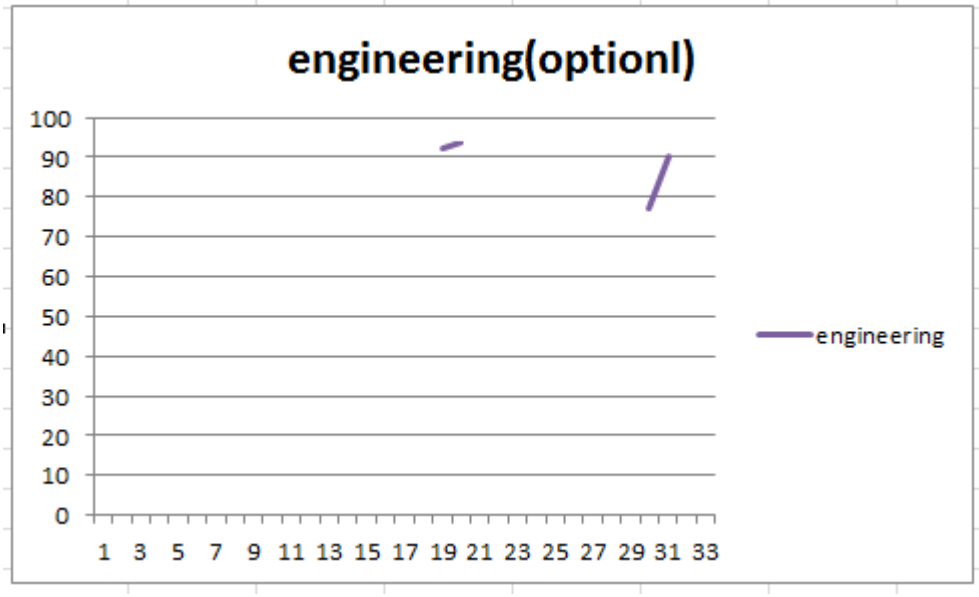
شكل(٥.١١): مادة الكيمياء.

مادة الأحياء



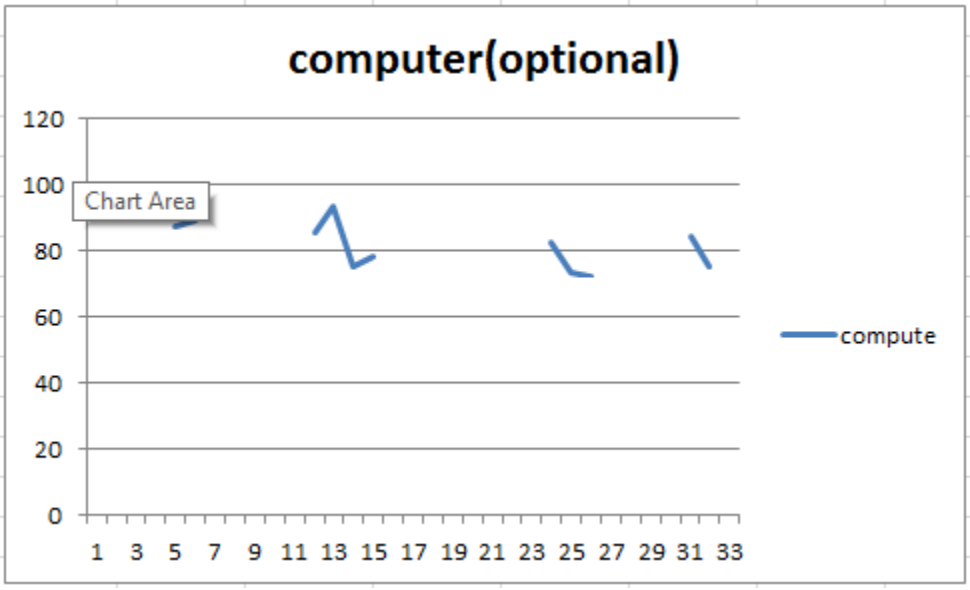
شكل(٥.١٢): مادة الأحياء.

مادة العلوم الهندسية



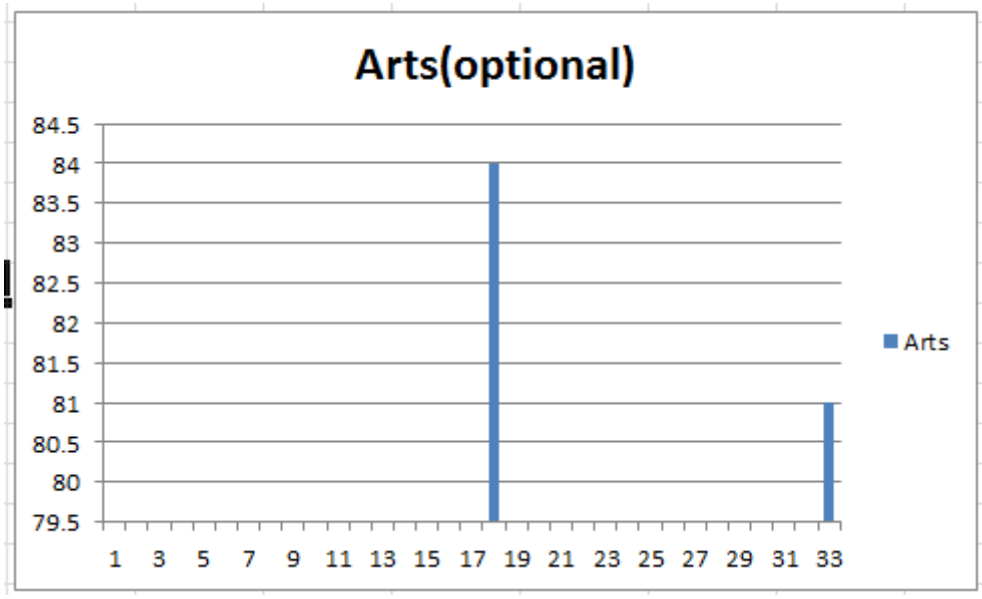
شكل(١٣.٥): مادة العلوم الهندسية.

مادة علوم الحاسوب



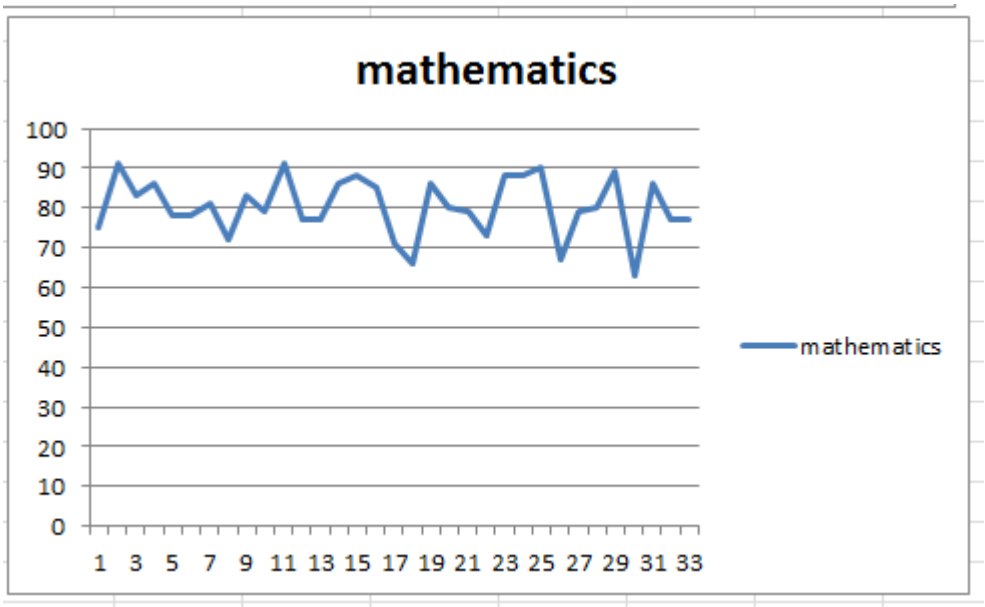
شكل(١٤.٥): مادة علوم الحاسوب.

مادة الفنون الجميلة



شكل(١٥.٥): مادة الفنون الجميلة.

مادة الرياضيات



شكل(١٦.٥): مادة الرياضيات.

4.5 التوصيات :

١- يجب تخزين البيانات بصورة جيدة لتكون في متناول اليد مما يساعد في إكمال مسيرة الأبحاث في هذا المجال.

٢- لزيادة كفاءة النتائج المتحصل عليها من هذا البحث نوصي بإضافة بيانات الطلاب الذين تعزر الحصول

علي بياناتهم نسبة لتواجدهم ضمن النظام القديم للقبول في الجامعة.

٣-البيانات الأساسية لكل طالب يجب الحصول عليها وتخزينها، بحيث لايكون هنالك مشاكل في جمعها عند الحوجة لتحليلها وإجراء التجارب عليها.

5.5 الخاتمة :

بعد الإطلاع على نتائج البحث وتحليلها ظهرت أهمية سجل الطالب الأكاديمي ومدى تأثيره علي المستويات الأكاديمية للطلاب قيد الدراسة ،وبالتالي ظهرت الحوجة إلي تحليل بيانات الطلاب للإستفادة منها في صنع القرارات ،حيث أن إجراء العديد من التجارب في هذا المجال يمكن المؤسسات التعليمية من تطوير الخطة المتبعة وزيادة الكفاءة .

وكما توضح من البحث والأبحاث السابقة أن تنقيب البيانات من الطرق الحديثة ذات الكفاءة العالية في تحليل وتجميع وتصنيف البيانات والتوصل إلى علاقات تزيد من إمكانية التحكم في النظام وتقديم أفضل ما يمكن للوصول للأهداف المرجوة.

6.5 المراجع :

[1] DATA MINING .THIRD EDITION .

[2] S. A. Kumar, “Mining Of Student Academic Evaluation Records in Higher Education,” pp. 67–70, 2012.

[3] V. Kumar and A. Chadha, “Mining Association Rules in Student ’ s Assessment Data,” vol. 9, no. 5, pp. 211–216, 2012.

[4] C. Romero, J. R. Romero, J. M. Luna, and S. Ventura, “Mining Rare Association Rules from e-Learning Data,” pp. 171–180.

[5] “Kovacic ~ Early prediction of student success.pdf.” .