

ministry of higher Education

Sabah University

College of science

Computer Denartment



وزارة التعليم العالي

جامعة صباح

كلية العلوم

قسم الحاسوب

تطبيق تقنيات التنقيب في البيانات لتقييم اداء طلاب

قسم الحاسوب - كلية العلوم

بحث لاستكمال متطلبات نيل درجة البكالوريوس في علوم الحاسب الالي

إعداد

الطالبة : سميره محمد علي القدم

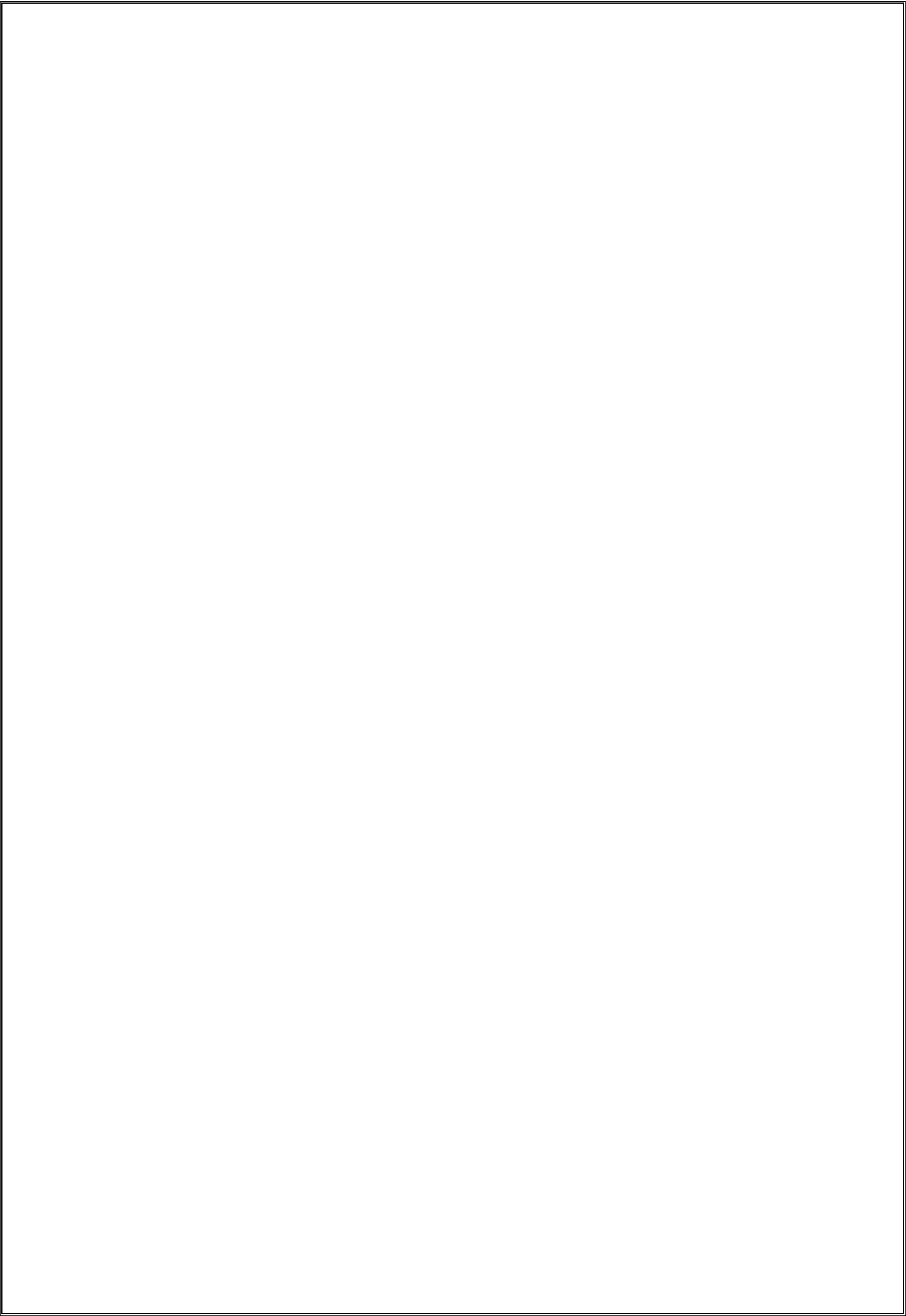
رقم القيد: 02120144

إشراف

أ. محمود حفص الدين

الشهر - العام

يناير - 2018



1. تمهيد

باستخدام تطبيقات الحاسوب الحديثة والتي تعتبر تكنولوجيا حديثة ذكية قائمة على جعل الحاسوب "يفكر كما يفكر ويفعل الإنسان" وهو ما يعرف بالذكاء الاصطناعي ، جاءت فكرة الكشف والتنقيب على هذه البيانات بطرق ذكية للمساعدة في حل المشاكل واتخاذ القرارات. تنقيب البيانات عملية تحليلية تمزج بين علم الذكاء الاصطناعي والإحصاء وتعليم الآلة وقواعد البيانات، وتعتبر خطوة من خطوات استكشاف المعرفة من قواعد البيانات.

1.1 مشكلة البحث

1. ضعف قدرة الوصول إلي تقويم شامل للمستوى الأكاديمي للطلاب خلال مرحلة دراسته.
2. صعوبة قياس نتائج التعلم وتحديد متطلباتها والعوامل التي تؤثر فيها.
3. النتائج المراد تقويمها قد تكون معنوية وتجريدية عامة.

1.2 أهمية البحث

1. توفير القدرة علي الاستكشاف والتركيز علي أهم المعلومات في سجل الطالب.
2. مساعدة في بناء التنبؤات المستقبلية مما يسمح بتقدير القرارات الصحيحة واتخاذها في الوقت المناسب.
3. مساهمة في تعديل معايير الأداء وزيادة أداء الأساتذة والجامعة بشكل عام.

3.1 دوافع اختيار البحث

1. تقييم الطلاب باستخدام تقنية التنقيب عن البيانات لتقويم وتحليل اداء الطالب
2. تطبيق ما تم دراسته

3.1 أهداف البحث

1-استخدام تقنية التنقيب في البيانات الطالب لتقويم وتحليل أداء الطلاب .

2-تحليل بيانات الطلبة قسم الحاسوب بكلية العلوم .

3-الهدف من التنقيب هو تحسين أداء الطالب أكاديميا..

4.1منهجية البحث

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

5.1حدود البحث

تشمل الدراسة تطبيق طرق تنقيب البيانات (Data mining) على سجل الطالب، وذلك باستخدام خوارزمية (Apriori) يتم تطبيقها على البيانات المأخوذة من استمارة سجل الطالب وهي عبارة عن درجات الطالب المتحصل عليها، وتشمل الطلاب الذين تم قبولهم سنة (2008 – 2009) وكان النظام التي تم التطبيق عليه نظام الساعات المعتمدة.

6.1هيكلية البحث

يتضمن البحث بالإضافة إلى هذا الباب الأبواب:-

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

الباب الثالث :ويتضمن الدراسات السابقة.

الباب الرابع :يتناول هذا الباب منهجية البحث. الباب الخامس:يتناول النتائج ومناقشتها.

2.تنقيب البيانات

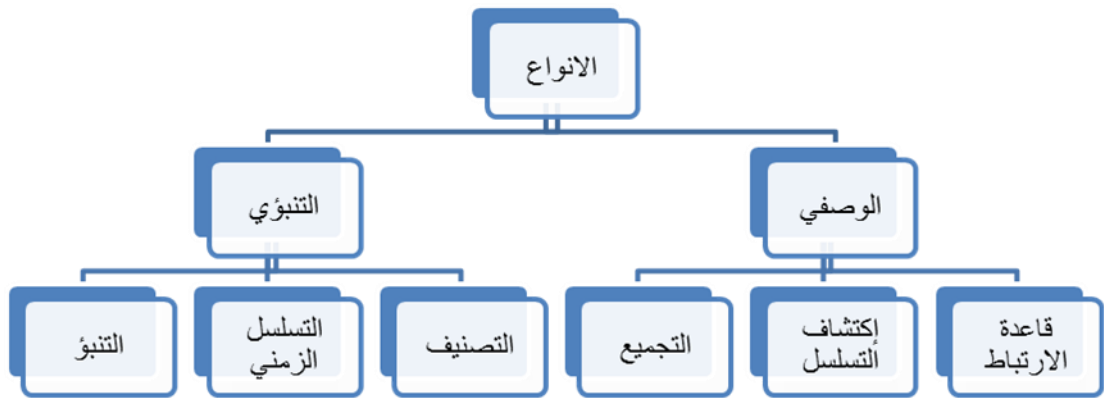
هي عملية بحث عن معرفة من البيانات دون [1] فرضيات مسبقة عما يمكن أن تكون هذه المعرفة . كما ويعرف التنقيب في البيانات على أنه عملية تحليل كمية بيانات عادةً ما تكون كمية كبيرة لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات.

يطلق أسم (نموذج) على العلاقات والبيانات المخصصة 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 لا تؤثر مطلقاً على طريقة تجميع البيانات ذاتياً. 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 استخدام الطرق الإحصائية العادية في تحليلها تعتبر عملية تنقيب البيانات عملية تحليلية والبحث في بيانات ضخمة وهائلة لاستخراج انماط مفيدة وإيجاد العلاقات ومدى الارتباط بين عناصرها.

1.2 أسباب تنقيب البيانات

1. إن التنقيب في البيانات يهدف إلى اكتشاف استخلاص انماط مفيدة ، وهي تكنولوجيا مفيدة أصبحت مهمة في ظل التطوير السريع وانتشار استخدام قواعد البيانات.
2. استخدامها يوفر للمؤسسات وأجهزة الأمن في جميع المجالات القدرة علي الاستكشاف والتركيز علي أهم المعلومات في قواعد البيانات.تركز تقنيات التنقيب علي بناء تنبؤات مستقبلية واستكشاف السلوك والاتجاهات ، مما يسمح بتقدير قرارات صحيحة واتخاذها في الوقت المناسب.
3. تجيب علي العديد من الأسئلة في وقت قياسي خاصة تلك النوعية من الأسئلة التي يصعب الإجابة عليها، إن لم يكن مستحيلا، باستخدام تقنيات الإحصاء الكلاسيكية والتي كانت إن وجدت تستغرق وقتا طويلا والعديد من إجراءات التحليل.

2.2 أنواع تنقيب البيانات



شكل (2.2) يوضح أنواع تنقيب البيانات

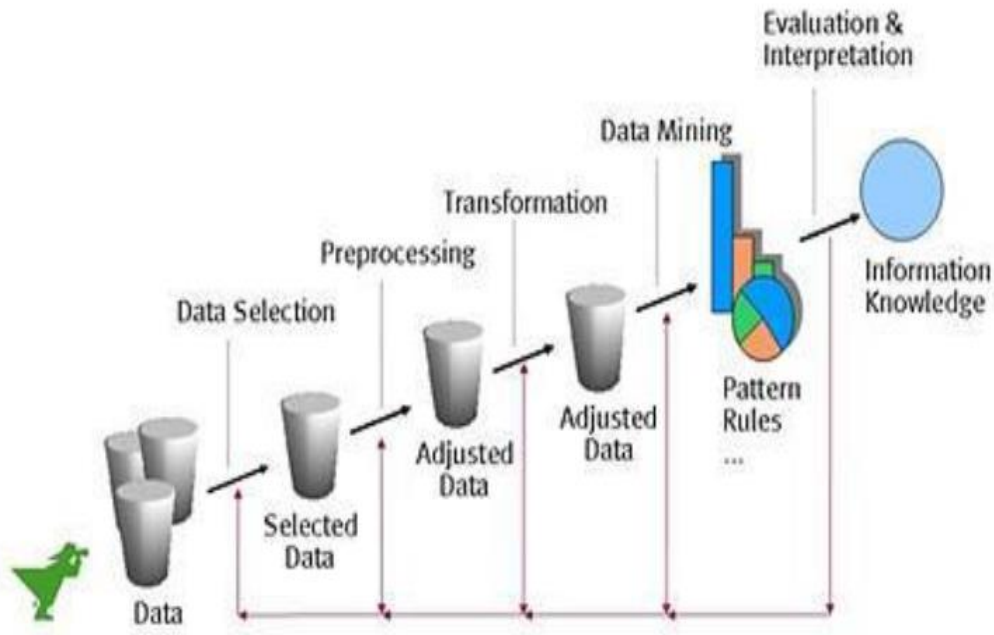
1.2.2 التنقيب الوصفي

يعتمد على إعادة تنظيم البيانات، والتنقيب في أعماقها لاستخراج النماذج الموجودة فيها كتشابه الزبائن الذي يسمح لك بإنشاء وصف بسيط عن مجموعة زبائن متشابهين ولا ستوجب وجود هدف لمثل تلك التشابه.

2.2.2 التنقيب التنبؤي

يحاول إيجاد أفضل التنبؤات اعتماداً من المعطيات كمعرفة المنتج الأفضل لزبون معين ، باختصار يعتمد هذا التنقيب على استخدام معلومات قديمة لتوقع ما سيحدث في المستقبل وتكون لدى مثل هذه البيانات هدف.

3.2 خطوات استكشاف المعرفة من قواعد البيانات (KDD) [9]



شكل (1.2) يوضح خطوات اكتشاف المعرفة من قواعد البيانات

1.3.2 اختيار البيانات (Data Selection)

في هذه المرحلة يتم تحديد واسترجاع البيانات الملائمة من قاعدة البيانات.

2.3.2 تهيئة البيانات (Data Preprocessing)

وهي مرحلة تجهيز وعزل البيانات المهمة والمفقودة أو التي تحتوي على شوائب من بقية البيانات كإلغاء المعلومات المتكررة، التصحيح الشكلي، معالجة البيانات الناقصة وجعلها جاهزة للتطبيق وتشمل هذه المرحلة (تنظيف البيانات، إزالة البيانات المفقودة، اشتقاق البيانات، دمج البيانات)

3.3.2 نقل البيانات (Data Transformation)

وهي عملية نقل البيانات التي تم اختيارها إلى شكل ملائم لإجراءات البحث والاسترجاع.

4.3.2 تنقيب البيانات (Data Mining)

في هذه المرحلة سيتم تطبيق أسلوب ذكي لاستخراج نماذج مفيدة بقدر الإمكان.

5.3.2 تقييم النموذج (Pattern Evaluation)

بعد استخراج النماذج المهمة التي تمثل المعرفة يتم تقييمها استنادا على مقاييس محددة.

6.3.2 تمثيل المعرفة (Knowledge Representation)

وهي المرحلة الأخيرة من مراحل اكتشاف المعرفة في قواعد البيانات والتي يراها المستفيد، وهي المرحلة الأساسية والتي تستخدم أسلوب مرئي لمساعدة المستفيد في فهم وتفسير النتائج المستخرجة.

4.2 مهام تنقيب البيانات

1.4.2 العقدة

الهدف منها هو تحديد الاتجاهات داخل البيانات ، ويحاول هذا الأسلوب العثور على مجموعات من العناصر التي توجد عادة معا .وهي عملية تقسيم البيانات إلى مجموعة [13] من الأصناف اعتمادا على اشتراكها بالخواص المتشابهة وأن العقدة هي تقسيم غير موجه للبيانات ، وهي عكس التصنيف ، كما أنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

2.4.2 اكتشاف التسلسل

تقنيات التسلسل تستخدم التحليل القائم على الوقت لانتزاع معلومات مفيدة ، وهي مماثلة للتجميع في إنها تستخدم لتحديد العناصر التي تحدث معا ، ولكن الأهم من ذلك أنها تستخدم لتحديد أي من العناصر يحدث أولا.

3.4.2 التصنيف

يستخدم التصنيف بشكل واسع في حل كثير من المشكلات من خلال تحليل مجموعة من البيانات ووضعها في شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية مثل: شراء سلع عبر الانترنت وهناك عدد من الطرق التي يمكن استخدامها في تصنيف [14]البيانات باستخدام الخوارزميات مثل الخوارزمية الإحصائية ، وخوارزميات الشبكات العصبية ، والعديد من خوارزميات تعلم الآلة .

4.4.2 الانحدار

هو تقنية تسمح بتحليل البيانات لوصف العلاقة بين متغيرين أو أكثر ، إن الانحدار يفترض أن توضع البيانات بنوع معروف من الدوال ومن ثم يتم تحديد أفضل دالة للبيانات المعطاة مثل: التنبؤ بأسعار الاسهم والاتجاهات.

5.4.2 التلخيص

توفر عرضا محكما للبيانات يحتوي علي التصور الكامل وتوليد التقرير و من الأمثلة على أساليب التلخيص نذكر: المتوسطات، و المجاميع، والإحصائيات الوصفية التي تتضمن مقاييس النزعة المركزية مثل المتوسط الحسابي و الوسيط والمنوال، ومقاييس التشتت مثل الانحراف المعياري. و على الرغم من أن مقاييس التلخيص تعطي صورة كبيرة عن بعض التفاصيل ذات العلاقة فإنها غالبا ما تهمل تفاصيل أخرى ذات أهمية كبيرة تتعلق بسلوك المستهلك خصوصا.[3]

6.4.2 قاعدة الارتباط

هي أحد التقنيات الواعدة من تنقيب البيانات كأداة من أدوات اكتشاف المعرفة ولديها القدرة على تحليل كميات هائلة من البيانات، وهي تسمح بالتقاط كل القوانين الممكنة التي تشرح بعض الصفات الموجودة اعتمادا علي وجود الصفات الأخرى، بمعنى آخر أنها قواعد ارتباطية معينة بين مجموعة من البيانات في قواعد البيانات مثل: تحليل سلة لشراء سلع[14] .

1.5.2 مقاييس قواعد الارتباط

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

القياس الموضوعي يشمل التحليل الإحصائي للبيانات مثل support ، confidence .

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

$X \rightarrow Y$ حيث إن كل منها تمثل مجموعة البنود (item set)

Support 2.5.2

هو عبارة عن نسبة توضح جزء من المعاملات التي تحتوي على كل من x و y.

جدول (1.2) قياس Support

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

Confidence 2.6.2

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

جدول (2.2) قياس Confidence

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

حيث

TID

TID	Items	Given X→Y
-----	-------	-----------

تمثل transaction id.

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

1 مجموعة البنود: عندنا يقال AK- item set يعني إن تكون مجموعة مكونة من k عنصر.

2 مجموعة البنود المكررة: هي التي لديها الحد الأدنى من support.

3. مقدمة

قام الباحثون بإجراء العديد من الأبحاث في ما يخص ملاحظة أداء الطالب الجامعي منذ نهايات القرن الماضي وقد تم تحليل درجات الطالب إحصائياً، وفي الآونة الأخيرة تم استخدام تقنيات حديثة في تنقيب البيانات في عديد من الدراسات بواسطة اداة الذكاء الاصطناعي WEKA هو عبارة عن برنامج خاص بخوارزميات التنقيب عن البيانات وتعلم الآلة. كما يحتوي هذا البرنامج على أدوات قادرة على التعامل مع الأمور التالية:

pre-processing, classification, regression, clustering, association rules, and visualization

نسرد لكم هذه الدراسات وما ثم من استخدام تقنيات عليها:

1.3 استخدام تقنية التصنيف لتقييم سجل الطالب

قام الباحثين بدراسة في إحدى [5] جامعات الهند بعمل بحث باستخدام تقنية 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 الأربعة سنوات الدراسية، وكان الغرض من هذا البحث هو توقع أداء الطالب في الامتحانات وإظهار التحسن في أدائه الأكاديمي.

وتم إعداد البيانات (Preprocessing) علي مجموعة البيانات طبقاً لقواعد الجامعة والعلامات التي حصل عليها الطالب تم تحويلها إلى الشكل الاسمي (nominal) وذلك لغرض البحث ، وتم استخدام إحدى أدوات تنقيب البيانات (weka) .

وكانت العوامل المستخدمة في الخوارزميتين هي:

1-الكفاءة في توقع نتيجة الطالب .

2-القيم الايجابية الصحيحة والقيم الايجابية الخاطئة وقيم إل ROC

3-الزمن المأخوذ في كل خوارزمية لإنتاج الناتج .

تقنيات التصنيف المختلفة يمكن استخدامها لتوقع نتيجة 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 واستخدامها 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) لعدده أسباب وهي أن الخوارزمية الأولى أكثر دقة من الأخرى وتؤدي أفضل منها وتأخذ وقت اقل.

2.3أستخدم تقنية قوانين الارتباط

1.2.3دراسه أولى

البيانات التي تطلبت للدراسه في هذا البحث تم أخذها من جامعة مشهورة في منطقة (Haryana) في الهند[6] لتحسين أداء التعليم بتحليل البيانات واكتشاف العوامل التي تؤثر على النتائج الأكاديمية ودراسة أداء الطلاب لكي تزيد فرص نجاح الطلاب. وتم عمل هذا البحث علي مجموعة من الطلاب وكان عددهم 30 طالب ، وتمت عليها بعض خطوات تجهيز البيانات (Data Preprocessing) هي (التنظيف Cleaning)التحويل (Transformation) وتم تحليل قوانين الارتباط باستخدام أحد أدوات تنقيب البيانات وهي (Tanagra) ، وذلك بتطبيق قوانين الارتباط 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 أو فشل الطلاب ، وباستخدام هذه التقنية تم التوصل إلى معرفة مدى اهتمام الطلاب بمواد معينة وتصميم المناهج من حيث طريقة التدريس وتقييمها لكي يتم قياس مدى تأثيرها على الطلاب الذين لم ينجحوا لكي يصل بهم إلى الحد المقنع من الأداء وكانت قيمة (confidence=16.667) وكانت قيمة (support=83.333) وكان

عدد (rules=2)، بالأخير تم الكشف عن عوامل مختلفة من حيث اهتمام الطالب ، تصميم المناهج ، تقييم المنهجيات التي تؤثر علي الطلاب وتسببت في إخفاقهم .

2.2.3 دراسة ثانية

قام الباحثين بدراسه في [7] جامعة Córdoba في إسبانيا وتطبيقا على البيانات التربوية في قسم علوم الحاسوب وتم تجميع هذه البيانات بواسطة نظام (Moodle) ووضعت في الجامعة ، وتم استخدام خوارزمية في قاعدة التجميع بهدف قياس الأداء واكتشاف العلاقات بين متغيرات البيانات وقد تم تطبيقها علي مجموعة مكونة من 230 طالب علي مدار 5 كورسات من قسم علوم الحاسوب وهذه البيانات تتضمن مجموعة الدرجات التي تحصل عليها الطالب في كل من (assignment, quizzes and forums) كل من خوارزميات RARM و ARM وكان الهدف من الدراسة هو اختبار الأداء والفائدة من تطبيق RARM علي مجموعة البيانات التعليمية.

وتم تطبيق تقنية قوانين الارتباط واستخدام خوارزمية (Apriori) وتطبيق أربعة أنواع منها:

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

وكانت المقارنة بين ARM و RARM كالآتي:

جدول (1.3) يوضح المقارنة بين مقترحات ARM و RARM .

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

وكانت النتائج المتحصل عليها هي:

جدول (2.3) يوضح القواعد المستخلصة باستخدام خوارزمية Apriori-Frequent .

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

1-من القاعدة الأولى نجد إن إذا كان عدد(Quizzes) كبير وعدد (Quizzes) التي نجح فيها الطالب كبير يحرز درجة الامتياز في الامتحان النهائي.

2-من القاعدة الثانية نجد إن الطالب الذي يقضي وقت طويل في الواجبات يحرز درجة الامتياز.

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

4-القواعد الثلاث الأخيرة خاصة بالطلاب الغائبين عن الامتحان النهائي ، حيث أنه إي طالب لم تكن له درجات في الواجبات ، الاختبارات ومشاركات المنتدى لم يجلس للامتحان النهائي.

3.3 استخدام تقنية التنبؤ

في هذا البحث نقوم بالتنبؤ المبكر (Early Prediction) من نتيجة [10]الدراسة لتي تستعمل بيانات تسجيل الطالب ، والهدف من ذلك البحث هو اكتشاف العوامل التي قد تؤثر على نتائج الدراسة في نظم المعلومات في كلية الفنون التطبيقية المفتوحة في منطقة نيوزلندا ، وكان عدد البيانات المستخدمة 453 طالب.

تم استخدام بيانات القيد (Enrolment Data) لتحقيق الأهداف الآتية:

- 1-بناء نماذج للتنبؤ المبكر من نتائج الدراسة باستخدام بيانات القيد.
- 2-تقييم النماذج باستخدام (Cross-Validation) و (Misclassification) وذلك لاتخاذ القرار.
- 3-النتائج الحالية يسهل فهمها من قبل المستخدمين.

كانت بيانات القيد تخص المعلومات الآتية:

الجنس ، تاريخ الميلاد، الأصول العرقية والعجز، حالة العمل ، تأهيل المدارس الثانوية الأكاديمية، Programming course ، Collage course ، course block ، تفاصيل الاتصال.يتم ملئ استمارة التسجيل في integrator يتم تسجيل تاريخ التسجيل ويصبح الطالب مسجل في(. course) وكانت تقنيات تنقيب البيانات المستخدمة هي التصنيف واختيار ميزة (feature selection)، وتم استخدام أربع أنواع من أشجار تصنيف وهي:

1-CHAID .

-2 Exhaustive CHAID .

-3 QUEST .

-4 CART .

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

1-الأصول العرقية.

2-Programming Course.

3-Course block.

ولكن لسوء أُلحظ أشجار التصنيف لم تعطى دقة عالية في النتائج ، وكان دقة شجرة CHAID هي 59.4% وكانت دقة شجرة CART هي 60.5%، وهي تقترح أن المعلومات المساعدة (الجنس ،العُمر ، الإِنتماء العرقي ، العجز ، المدرسة الثانوية ، حالة العملِ ، والتسجيل المبكر) تجمع أثناء عملية التسجيلِ ، ولا تحتوي على معلومات كافية للفصل بين الناجحين والفاشلين.

تهتم هذه الدراسة بالنواحي الاجتماعية والبيئة الدراسيه التي تؤثر علي الطالب ، بمعنى آخر من خلال بيانات التسجيل نستطيع التحكم في مدى نجاح أو فشل الطالب.

4. تمهيد

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

1.4 أولا اختيار البيانات

تم الحصول على البيانات من قبل قسم علوم الحاسوب ، وتتمثل هذه البيانات في نتائج الطالب التي تحصل عليها خلال دراسته الجامعية على مدار الأربع سنوات . وقد تحصلنا في بداية تجميع البيانات علي 97 طالب وطالبة من ولكن بعد المعالجات التي تمت عليها قل هذا العدد إلى أن تم التعامل مع 36 طالب وطالبة.

2.4 ثانيا تهيئة البيانات

بعد اكتمال جمع البيانات من مصادرها واستبعاد البيانات الغير مكتملة البيانات من مجموعة البيانات لأن هذه البيانات لابد أن تكون مكتملة لأنها تحتوي على درجات طلاب وهذه الدرجات لايمكن تعويضها بشيء آخر لأنها لابد أن تكون حقيقة . وكانت مكتوبة في شكل (Excel sheet) وكان الملف يحتوي على جميع المواد خلال الأربع سنين ، وكل ماده تحتوي على جميع التفاصيل للمادة العلمي والنظري وأعمال السنة والنتيجة النهائية والتقدير، حيث تم فصل الطلاب الذين تعذر الحصول علي درجاتهم الخاصة بنتائج، بالإضافة إلي تهيئة الشكل العام للملف بحيث يناسب أداة تنقيب البيانات المستخدمة وهي (WEKA) وهي تقبل البيانات في صورة ملف (ARFF) أو قاعدة بيانات.

3.4 ثالثا تنقيب البيانات

في هذه المرحلة تم استخدام تقنية قاعدة الارتباط وتطبيق خوارزمية (Apriori) باستخدام برنامج (Weka).

4.4 ما تحتويه Weka

يعدنا برنامج (Weka) [3] بمجموعة من خوارزميات تعليم 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: الانحدار 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 ملف (ARFF) للعلاقة: (project)

@relation project

@attribute Arc numeric

@attribute ENC numeric

@attribute physics numeric

@attribute Mathematics numeric

@attribute compute numeric

@attribute Statistics numeric

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

في واجهة (Weka) التفاعلية يمكن تحديد طريقة التعلم المراد استخدامها من القائمة، والكثير من الطرق لديها مجموعة من المعاملات الانضباطية التي يمكنك الوصول اليها من خلال ورقة الملكية أو من خلال محرر كائن، ويتم استخدام 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 استخدام (Weka)

أسهل الطرق لاستخدام (Weka) هو من خلال استخدام واجهة المستخدم الرسومية وتسمى المستكشف (Explorer)، وهذا الأخير يتيح الوصول الى جميع منشآت (Weka) بإستخدام إختيار القائمة وملاً الإستمارة مثلاً 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

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 الاختيار من بين هذه الأربع واجهات.

8.4 البداية

افترض أن لديك مجموعة من البيانات وتريد بناء شجرة القرارات منها ، أولاً لابد من تهيئة هذه البيانات وإطلاق المستكشف والقيام بتحميل البيانات فيه ، ثانياً لابد من اختيار طريقة بناء شجرة القرار وبناء الشجرة وتفسير المخرجات في المستكشف 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)

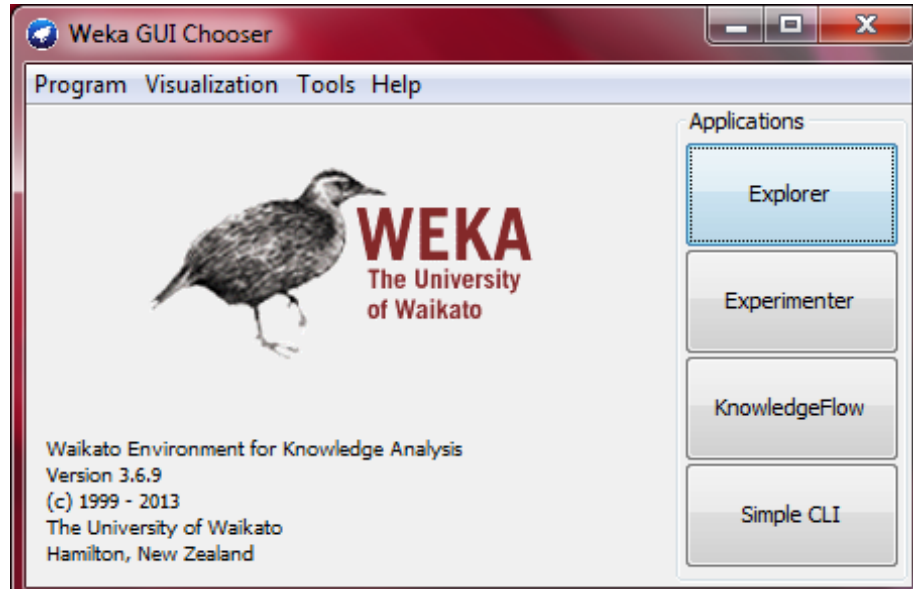
غالبا ما يتم عرض البيانات في صورة (Excel Sheet) أو قاعدة بيانات ، على أي حال الطريقة الأصلية للتخزين في (Weka) هي في صورة (ARFF) ويمكن بكل سهولة التحويل من صورة برنامج الجدولة الى صورة (ARFF) معظم ملفات (ARFF) تتكون من قائمة من الحالات وقيم خواص لكل من الحالات التي تم الفصل بينها بالفاصلة كما ذكرنا سابقا.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

A1	Q	P	O	N	M	L	K	J	I	H	G	F	E	D	C	B	A
aut2008	d	EN101	aut2008	d	ARAB1	aut2008	c	CS101	sp2009	d	ST101	aut2008	d	PH101	sp2008	d	MA101
aut2008	d	EN101	sp2009	d	ARAB1	sp2009	c	CS101	sp2011	c	ST101	sp2011	c	PH101	sp2009	d	MA101
aut2008	b	EN101	aut2011	b	ARAB1	aut2008	d	CS101	aut2013	d	ST101	sp2009	c	PH101	sp2009	c	MA101
aut2008	d	EN101	aut2008	d	ARAB1	aut2008	d	CS101	aut2009	d	ST101	sp2010	d	PH101	aut2010	c	MA101
aut2008	d	EN101	sp2009	d	ARAB1	aut2008	d	CS101	sp2011	c	ST101	aut2012	b	PH101	sp2009	d	MA101
aut2008	c	EN101	sp2009	d	ARAB1	aut2008	d	CS101	sp2011	c	ST101	aut2012	b	PH101	sp2009	d	MA101
aut2008	a	EN101	aut2008	d	ARAB1	sp2009	d	CS101	sp2011	d	ST101	aut2008	a	PH101	sp2009	d	MA101
aut2008	a	EN101	aut2008	d	ARAB1	aut2008	c	CS101	sp2009	d	ST101	aut2008	c	PH101	sp2009	b	MA101
aut2008	d	EN101	aut2008	c	ARAB1	sp2009	c	CS101	aut2010	d	ST101	aut2008	d	PH101	sp2009	d	MA101
aut2008	d	EN101	sp2009	d	ARAB1	sp2009	c	CS101	aut2011	b	ST101	sp2011	c	PH101	sp2009	d	MA101
aut2008	d	EN101	sp2009	d	ARAB1	sp2009	d	CS101	aut2010	c	ST101	aut2012	d	PH101	sp2009	b	MA101
aut2008	d	EN101	aut2008	d	ARAB1	aut2008	d	CS101	sp2010	d	ST101	sp2009	c	PH101	sp2009	c	MA101
aut2008	a	EN101	aut2008	b	ARAB1	aut2008	c	CS101	sp2009	d	ST101	aut2008	d	PH101	sp2009	d	MA101
aut2008	c	EN101	aut2008	d	ARAB1	aut2008	d	CS101	aut2010	d	ST101	aut2008	d	PH101	sp2009	d	MA101
aut2008	c	EN101	aut2008	d	ARAB1	sp2009	c	CS101	aut2011	b	ST101	sp2009	d	PH101	sp2009	d	MA101
aut2008	b	EN101	aut2008	d	ARAB1	sp2009	d	CS101	aut2011	a	ST101	sp2011	c	PH101	sp2009	d	MA101
sp2012	d	EN101	sp2012	c	ARAB1	sp2012	d	CS101	sp2012	d	ST101	sp2012	c	PH101	sp2012	c	MA101
sp2012	c	EN101	sp2012	a	ARAB1	sp2012	d	CS101	sp2012	b	ST101	sp2012	c	PH101	sp2012	c	MA101
sp2012	c	EN101	sp2012	d	ARAB1	sp2012	c	CS101	sp2012	b	ST101	sp2012	b	PH101	sp2012	b	MA101
sp2012	c	EN101	sp2012	b+	ARAB1	sp2012	d	CS101	sp2012	c	ST101	sp2012	d	PH101	sp2012	d	MA101
aut2009	d	EN101	sp2010	c	ARAB1	sp2010	d	CS101	sp2010	d	ST101	aut2010	c	PH101	aut2010	c	MA101
aut2009	d	EN101	aut2009	c	ARAB1	aut2009	d	CS101	sp2010	d	ST101	aut2010	d	PH101	aut2010	d	MA101
aut2010	d	EN101	aut2010	b+	ARAB1	aut2010	c	CS101	aut2010	b+	ST101	aut2010	c	PH101	aut2010	d	MA101
aut2011	b	EN101	aut2010	a	ARAB1	aut2010	c	CS101	aut2010	c	ST101	aut2010	c	PH101	aut2010	d	MA101
sp2011	b	EN101	aut2010	b+	ARAB1	aut2010	d	CS101	aut2010	d	ST101	aut2010	d	PH101	sp2011	a	MA101
sp2011	b	EN101	aut2010	b	ARAB1	aut2010	d	CS101	sp2011	c	ST101	sp2011	c	PH101	aut2010	d	MA101
sp2010	d	EN101	aut2010	c	ARAB1	aut2010	d	CS101	aut2010	d	ST101	sp2011	b+	PH101	sp2011	c	MA101

الشكل: (4. 1) يوضح ترتيب البيانات في جدول اكسل..

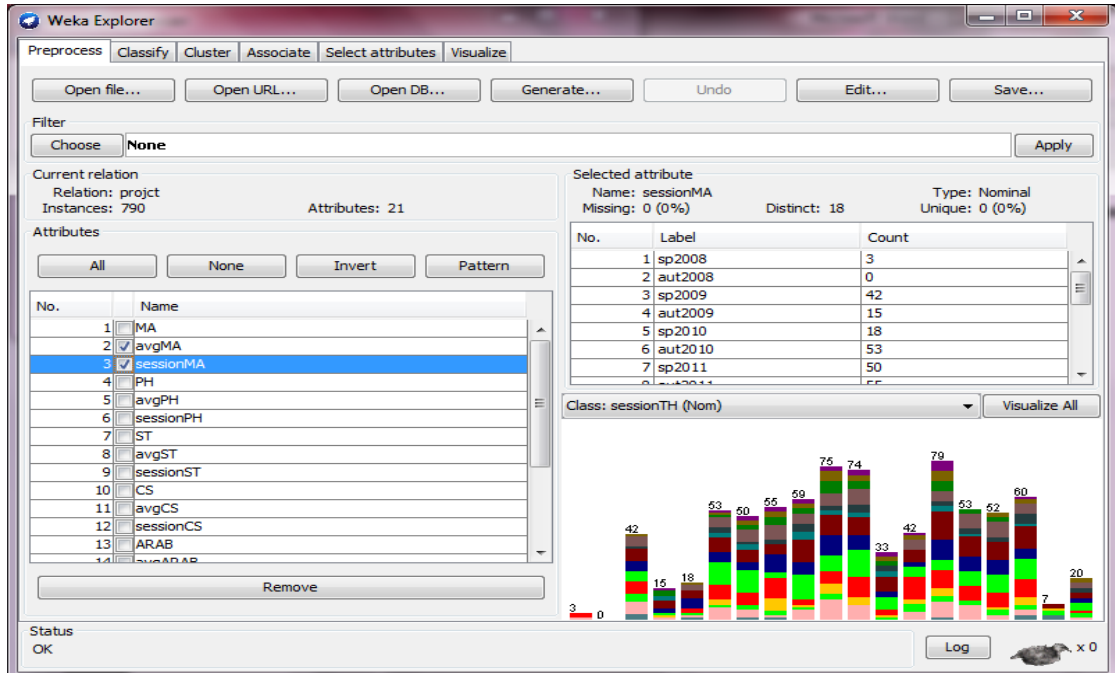
10.4 تحميل البيانات في المستكشف

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



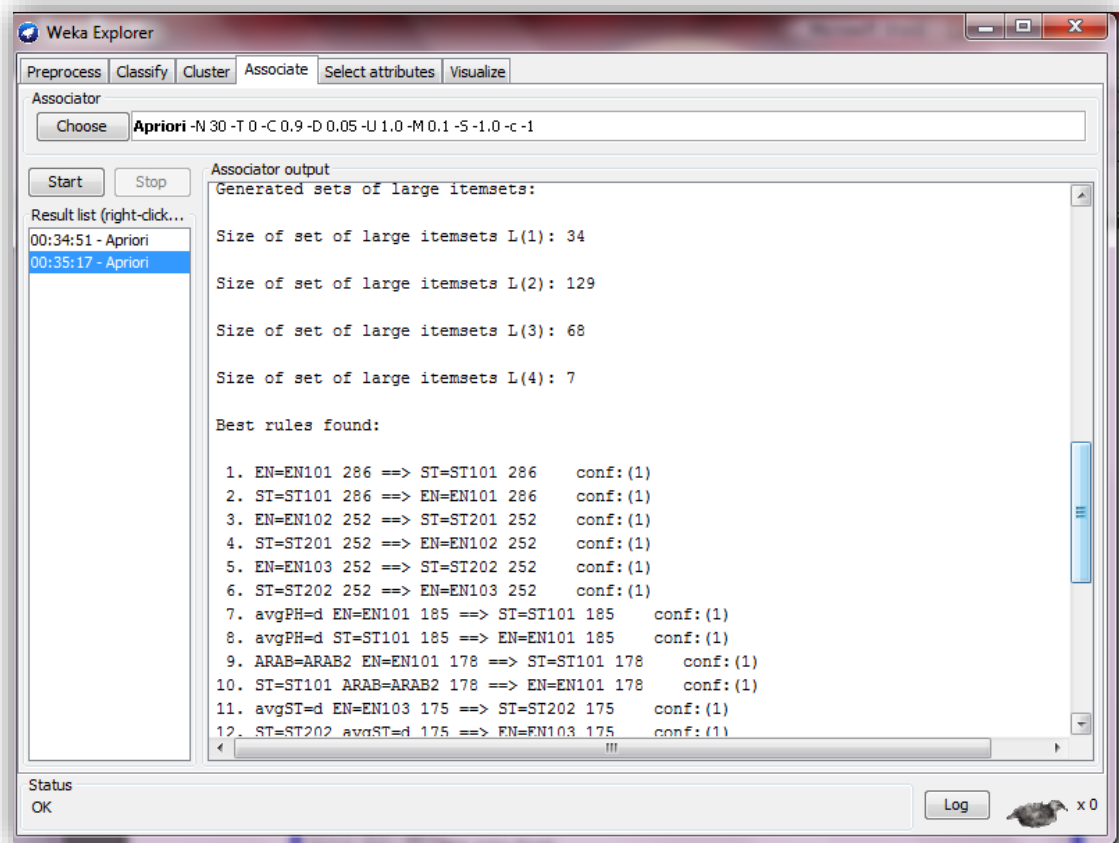
الشكل : (4. 2) واجهة الاختيار الرسومية .

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



الشكل : (3.4) 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 اجراء عملية التجميع عليها ويوجد أيضا خيار إختيارها جميعها واجراء التجميع عليها .ومن ثم بالرجوع للست قوائم الأولى نقوم بإختيار start ثم associate ، بعدها يقوم البرنامج بعمل توليد للنتائج والمخرجات كما هو موضح في الشكل: (4.4)



الشكل (4.4) شكل الشاشة عندما يتم الحصول على النتائج.

11.4 خوارزمية Apriori

هي خوارزمية للتنقيب عن [2] العناصر المتكررة وتطبيق قوانين الارتباط على (transactional databases). وتعمل عن طريق تحديد العناصر الفردية المتكررة في قاعده البيانات وتوسيعا لتشمل مجموعات أكبر. وأكبر بند هو البند الأكثر ظهورًا في قاعدة البيانات، ومجموعات البند المتكررة (The frequent item sets) (يتم تحديدها عن طريق خوارزمية Apriori) التي تستخدم لتحديد قوانين الارتباط هذه الخوارزمية مصممة للعمل على قواعد البيانات التي تحتوي على (transactions)، وتم تصميم الخوارزميات الأخرى

لإيجاد قوانين الارتباط في حالة عدم وجود (transaction) أو عدم وجود (timestamps) وينظر إلى كل (transaction) أنها مجموعة من الوحدات (item set)، تستخدم الخوارزمية منهج "bottom up" حيث يتم تمديه هذه المجموعات المرشحة وتنتهي هذه الخوارزمية عندما يتم العثور على أي تمديدات أخرى ناجحة. وتستخدم هذه 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: [2]

```

Apriori(T, ε)
  L1 ← {large 1 – itemsets}
  k ← 2
  while Lk-1 ≠ ∅
    Ck ← {a ∪ {b} | a ∈ Lk-1 ∧ b ∈ ∪ Lk-1 ∧ b ∉ a}
    for transactions t ∈ T
      Ct ← {c | c ∈ Ck ∧ c ⊆ t}
      for candidates c ∈ Ct
        count[c] ← count[c] + 1
    Lk ← {c | c ∈ Ck ∧ count[c] ≥ ε}
    k ← k + 1
  return ∪k Lk

```

توضيح الرموز

1.(T)توضح Transaction.

2.(€)توضح support .

3.(CK) توضح المجموعة المرشحة في المستوى K .

4.(Count[c])توضح حقل هيكلية البيانات التي تمثل المجموعات المرشحة c ،الذي من المفترض أن تكون قيمته الابتدائية تساوي صفرا.

5.(L)توضح اكبر بند (Large item set)

1.5 مقدمة

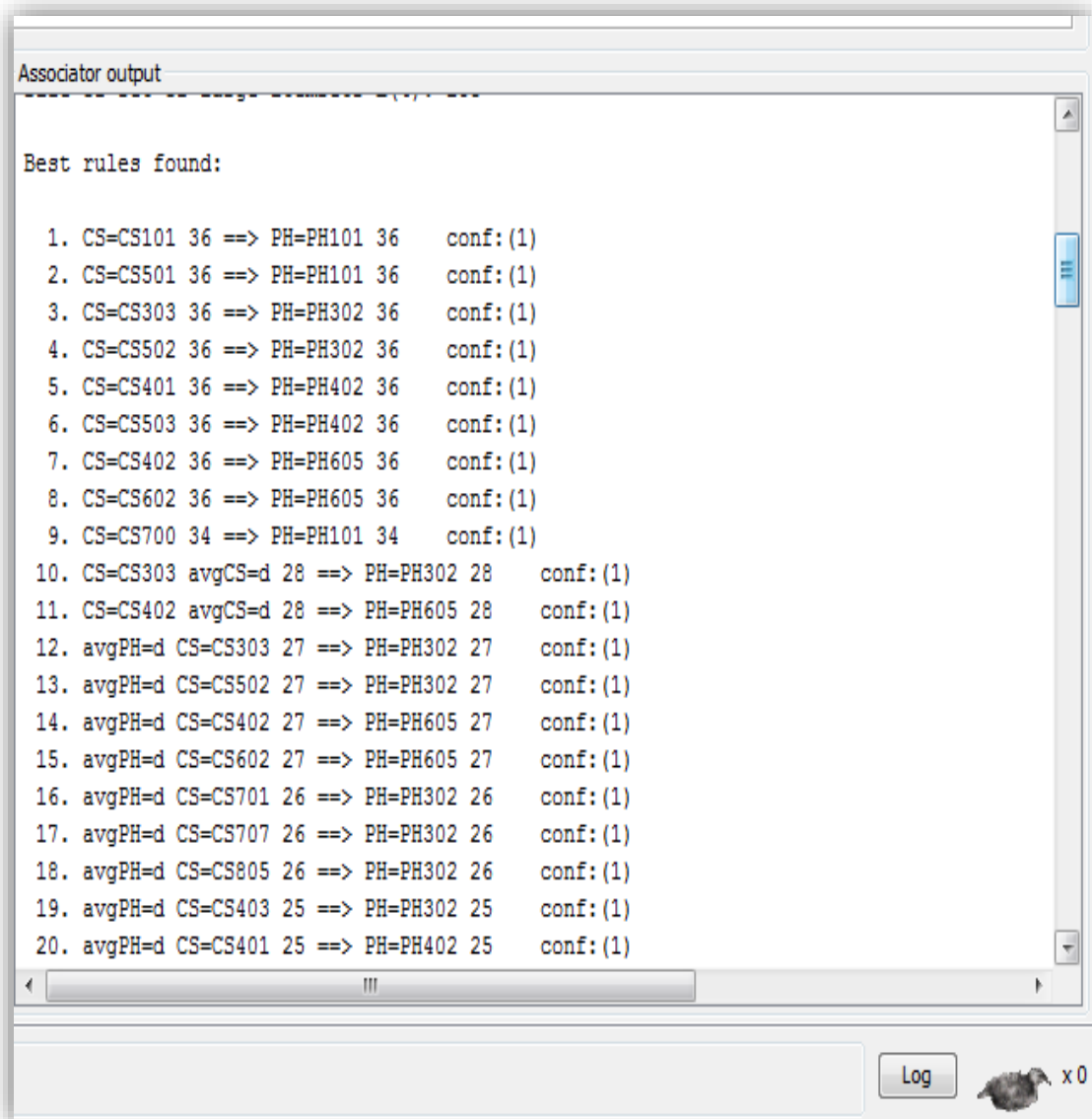
بعد اجراء العديد من التجارب على برنامج (Weka) وذلك بتغيير قيم (Confidence) و (Support) وتغيير الصفات التي يتم إختيارها في مرحلة التهيئة للبيانات (Preprocessing) تم التحصل على كمية كبيرة من القواعد (Rules) وبعد تحليل تلك القواعد وجد أن بعضها يصنف عمى أنه بديهي (Trivial) والبعض الآخر تتباين درجة منطقيته على أساس قيمة (Confidence) بعد تحليل النتائج المتحصل عميها كافة قمنا بإختيار القواعد التي توضح قوة إرتباط المواد ببعضها من حيث التأثير على مستوى 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.

يوضح ملخص النتائج المتحصل عليها بعد تطبيق خوارزمية (Apriori) موضحا عدد

المجموعات المتكررة (Frequent item sets) التي تم تنقيحها وعدد القواعد المكتشفة ومتوسط كل من (confidence and support) لكل قاعدة:

جدول (1.5): ملخص النتائج المتحصل عليها

Confidence	Support	Set n .Rules	sets Large item	N .item
0.9	0.1	10	5	1
0.2	0.01	100	7	1
0.2	0.0010	200	6	2



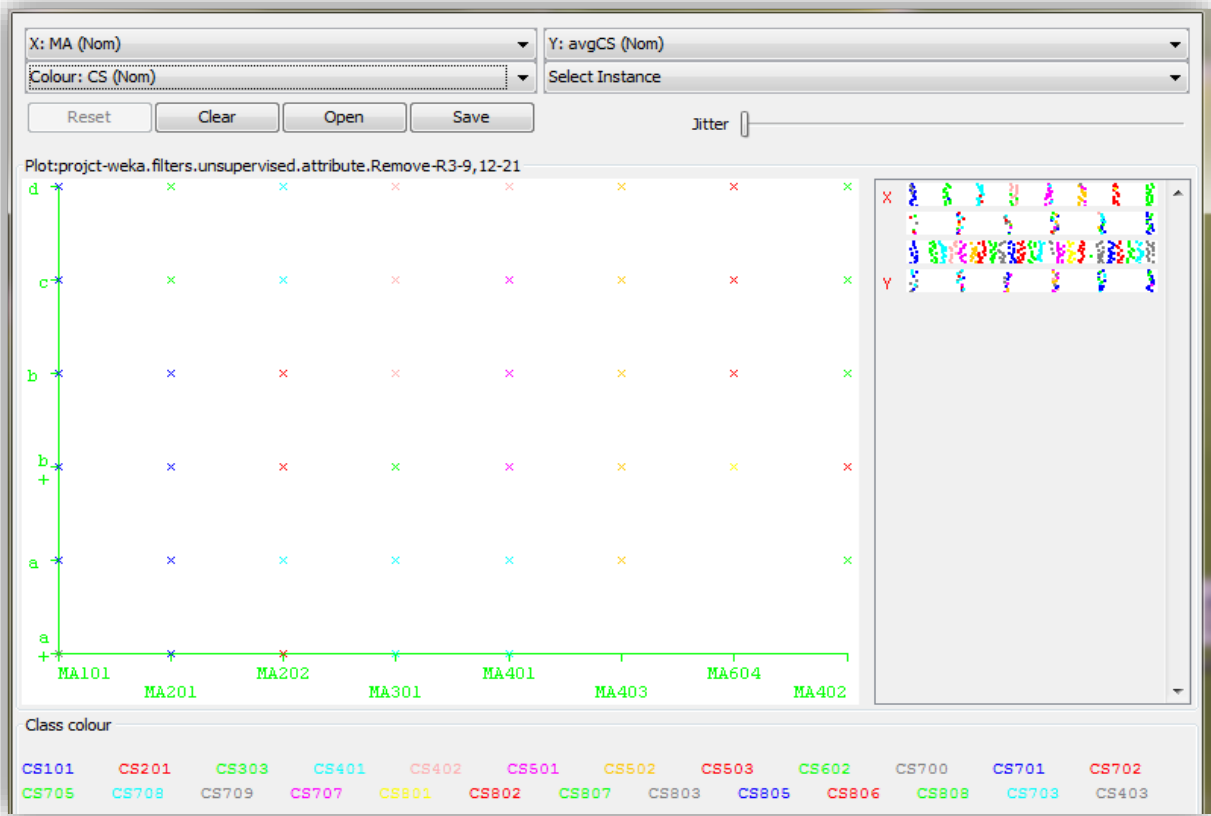
الشكل (1.5) جزء من النتائج المتحصل عليها

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 (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 من ثم بالرجوع للست قوائم الأولى نقوم بإختيار نمط من انماط التنقيب وهو (visualizr) توضح مرئية النتائج للمساعدة في تسهيل الملاحظات لمخرجات خوارزمية التنقيب وفهمها، وكانت القواعد المعتمدة كالآتي

اولا مواد الحاسوب مع مواد الرياضيات

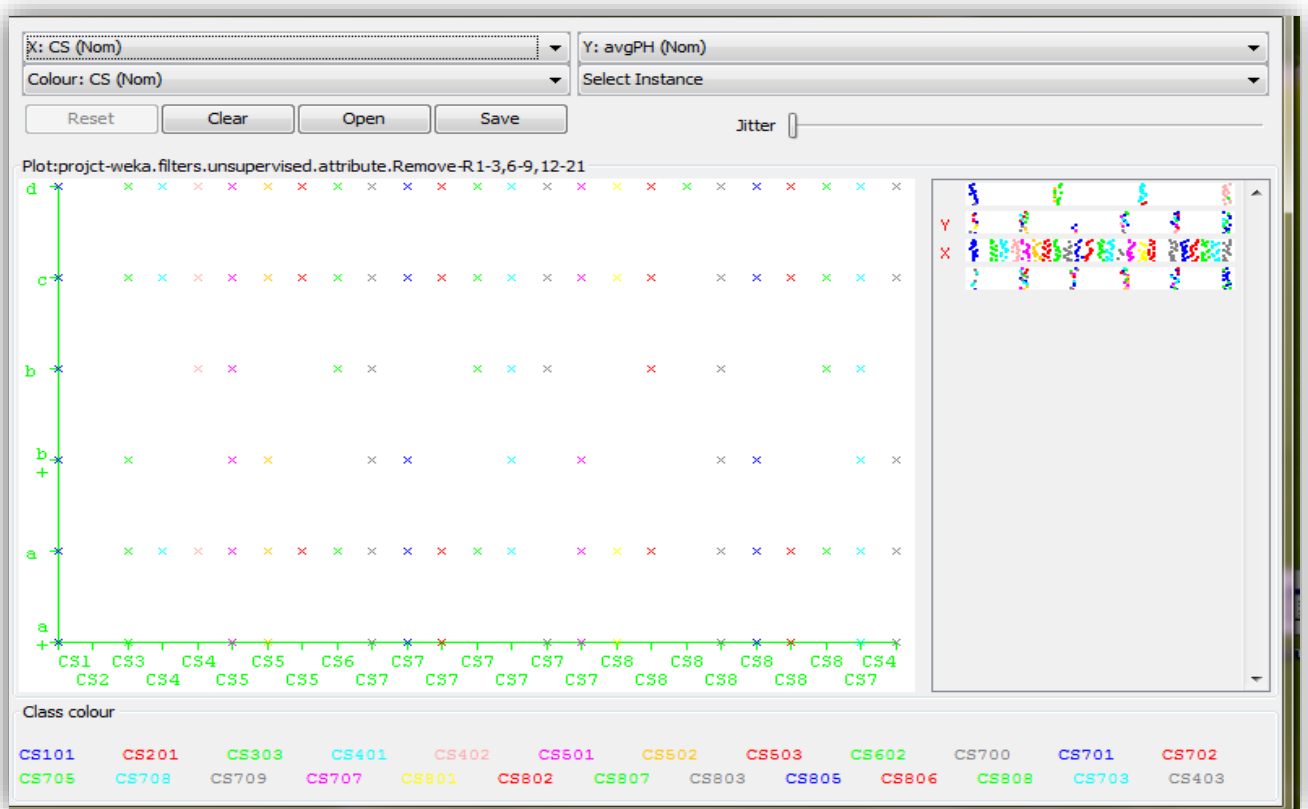
1. $\text{avgMA}=\text{d CS}=\text{CS30333} \implies \text{MA}=\text{MA20133}$
2. $\text{CS}=\text{CS401} \text{avgCS}=\text{c15} \implies \text{MA}=\text{MA202}$
3. $\text{avgMA}=\text{d CS}=\text{CS801} \text{avgCS}=\text{d 12} \implies \text{MA}=\text{MA604 12}$
4. $\text{MA}=\text{MA604} \text{avgMA}=\text{b 4} \implies \text{CS}=\text{CS700} \text{avgCS}=\text{b 6}$
5. $\text{avgMA}=\text{d CS}=\text{CS402} \text{avgCS}=\text{d 21} \implies \text{MA}=\text{MA301 21}$



الشكل (2.5) يوضح معدل الطلاب بين مواد الحاسوب والرياضيات

ثانيا مواد الحاسوب مع مواد الفيزياء

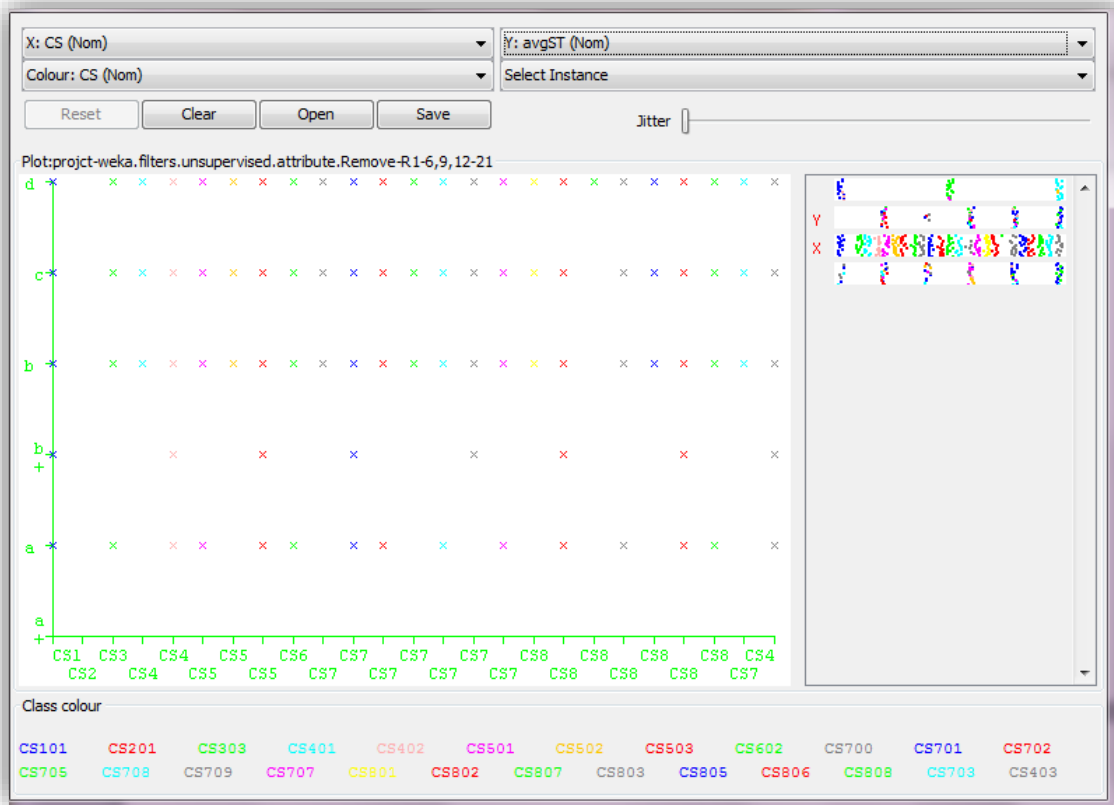
1. $\text{avgPH}=c$ $\text{CS}=\text{CS801}$ $\text{avgCS}=c$ 4 \implies $\text{PH}=\text{PH402}$ 4
2. $\text{avgPH}=a+$ \implies $\text{CS}=\text{CS703}$ 4 \implies $\text{PH}=\text{PH605}$ 4
3. $\text{avgPH}=a$ $\text{CS}=\text{CS101}$ $\text{avgCS}=c$ 4 \implies $\text{PH}=\text{PH101}$ 2
4. $\text{avgPH}=d$ $\text{CS}=\text{CS806}$ $\text{avgCS}=b+$ 3 \implies $\text{PH}=\text{PH402}$



الشكل (3.5) يوضح معدل الطلاب بين مواد الحاسوب والفيزياء

ثالثا مواد الحاسوب ومواد الاحصاء

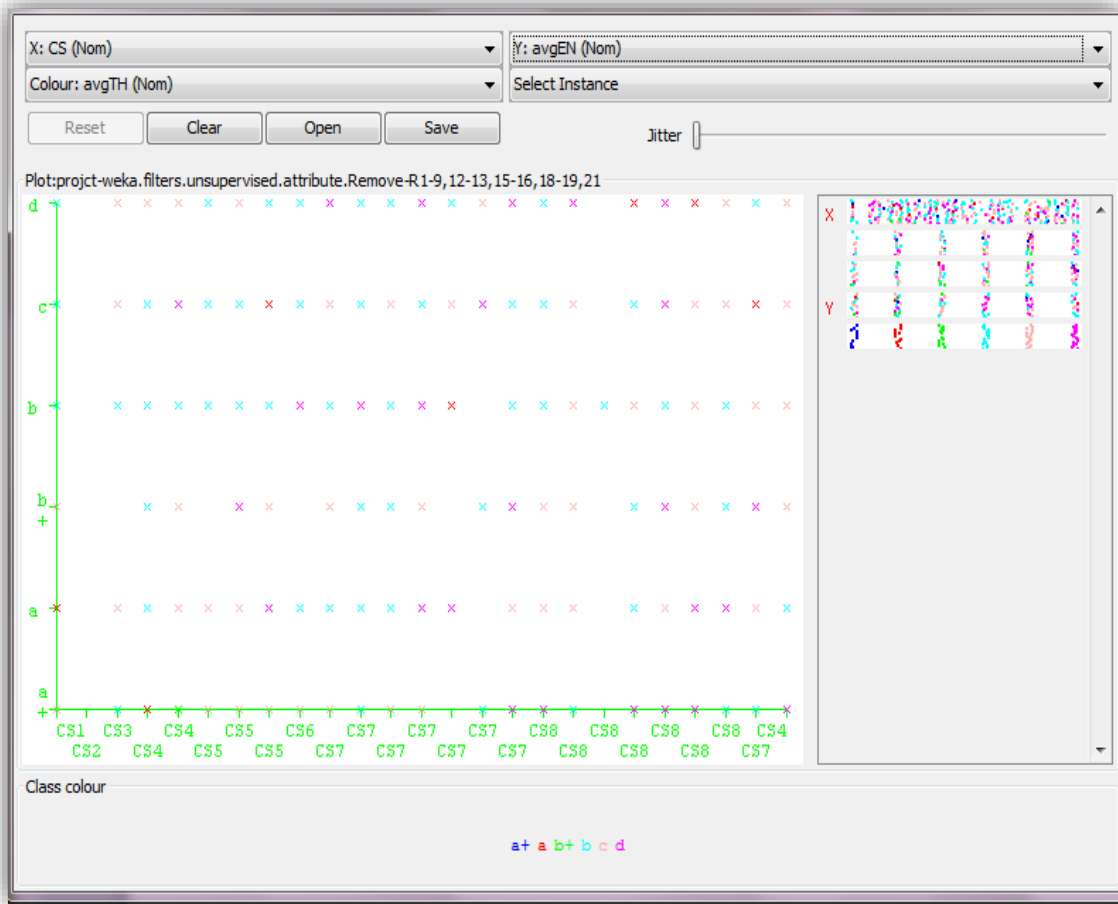
1. $CS=CS805$ $avgCS=a+3 \implies ST=ST202$ 3
- 2.. $avgST=c$ $CS=CS700$ $avgCS=c$ 4 $\implies ST=ST202$ 4
3. $avgST=c$ $CS=CS502$ $avgCS=b$ 3 $\implies ST=ST202$ 3
4. $avgST=c$ $CS=CS803$ $avgCS=b$ 5 $\implies ST=ST201$



الشكل (4.5) يوضح معدل الطلاب بين مواد الحاسوب والاحصاء

رابعاً مود الحاسوب مع المواد الثقافية

1. $CS=CS602 \text{ avgCS}=c \ 14 \implies ARAB=ARAB2 \ 14$
2. $CS=CS101 \text{ avgCS}=c \ 13 \implies ARAB=ARAB1$
3. $CS=CS806 \text{ avgARAB}=d \ 20 \implies ARAB=ARAB2 \ 20$
4. $CS=CS401 \text{ avgEN}=b+ \ 4 \implies EN=EN103 \ 4$
5. $CS=CS705 \text{ avgCS}=c \ \text{avgTH}=d \ 4 \implies TH=LIBH \ 4$



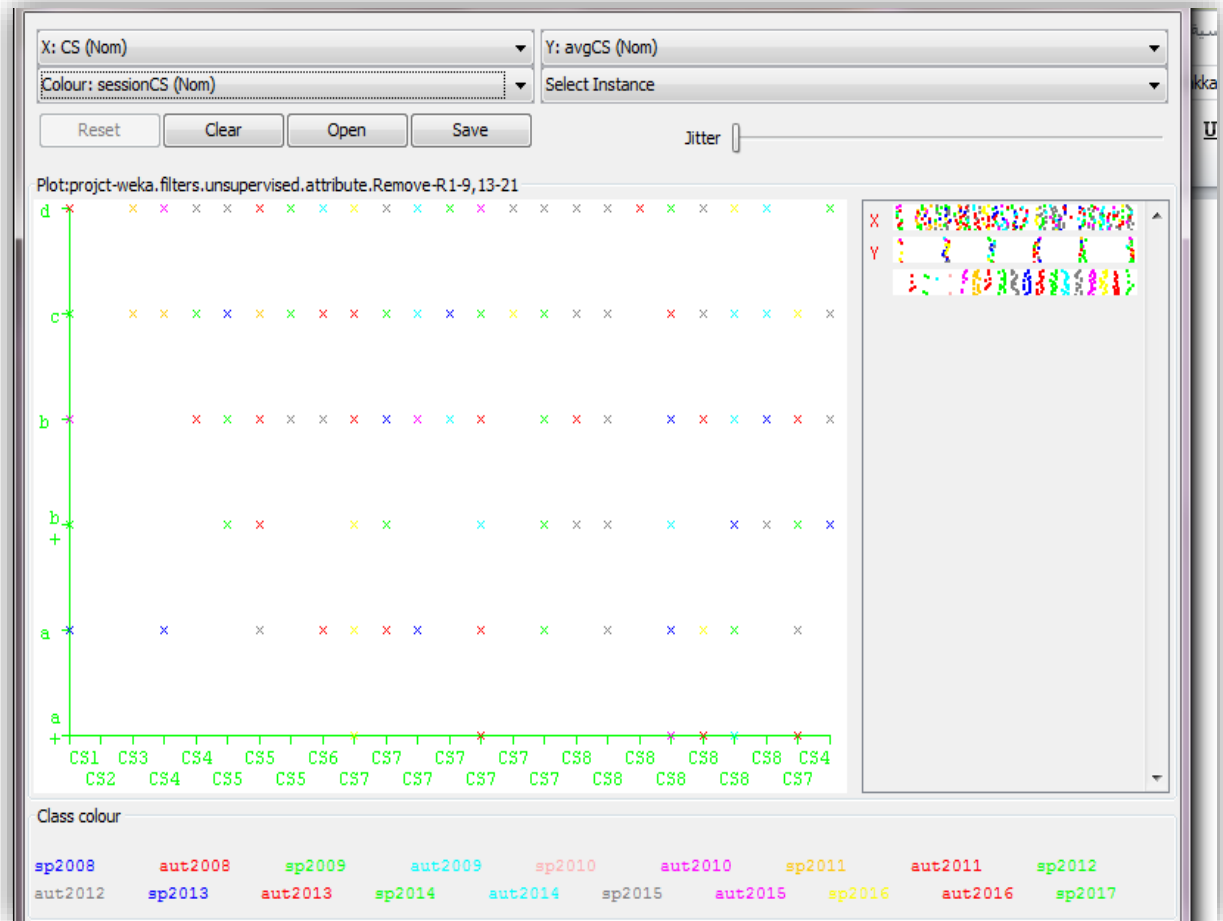
الشكل (5.5) يوضح معدل الطلاب بين مواد الحاسوب والفيزياء

رابعاً مواد الحاسوب مع بعضهم

1. CS=CS702 36 ==> avgCS=d 18 CS=CS801 avgCS=d 16
2. CS=CS702 ==> avgCS=b 3 CS=CS705 7 ==> avgCS=d 5
3. avgCS=b ==> CS=CS700 5 CS=CS707 ==> avgCS=a 5

4. avgCS=b+ 5 ==> CS=CS701 ==> CS=CS805 8 ==> avgCS=d

5. CS=CS803 3 ==> avgCS=c 2 ==> CS=CS802 ==> avgCS=d



الشكل (6.5) يوضح معدل الطلاب بين مواد الحاسوب والفصل الدراسي

2.5 مناقشة النتائج

المواد التي تدرسها الطالب مع بعضها تساعد في تحسين اداء الطلبة

جدول (2.5) يوضح ابرز المواد تساعد في تحسين اداء الطلبة

مواد المساندة	تقديره فيها	مواد التخصص	تقديره فيه
هندسة تحليلية	جيد	دوائر منطقية	جيد
الجبر الخطي	جيد	مقدمة طرق البحث	جيد جدا
الفيزياء عامة	ممتاز	مقدمة علم الحاسوب	ممتاز
الفيزياء الالكترونية	جيد	مشروع التخرج	ممتاز
الفيزياء المغناطيسية	جيد	انظمة التشغيل	جيد
الاحصاء احتمالات	ممتاز	امن الشبكات	ممتاز
الاحصاء احتمالات	جيد	مقدمة طرق البحث	جيد
الاحصاء العامة 2	جيد جدا	الوسائط المتعددة	جيد
مادة الحاسوب طرق البحث	جيد جدا	شبكات الحاسوب	ممتاز

بعض المواد التي درسها الطالب تبين فيها ضعف في اداءه

جدول(3.5)يوضح ابرز المواد تبين فيها تدني اداء الطلبة

مواد المساندة	تقديره فيها	مواد التخصص	تقديره فيه
الجبر الخطي	مقبول	انظمة التشغيل	مقبول
متسلسلات ومتتبعات	مقبول	امن الشبكات	جيد
رياضيات 3	مقبول	لغة التجميع	مقبول
فيزياء مغناطيسية	مقبول	برمجة قواعد البيانات	مقبول
برمجة انترنت	مقبول	انظمة تشغيل	مقبول
تحليل نظم	جيد جدا	هندسة برمجيات	مقبول

كذلك يمكننا استخدام التقنيات الأخرى في تقييم النتائج::

1.2.5 تقنية التصنيف

التصنيف مهمة تعليم دالة معينة F لربط [8] مجموعة الخصائص (X) بالفئة (Class) معرفة مسبقا تسمى (y) تعرف الدالة f أيضا بنموذج التصنيف (classification model).

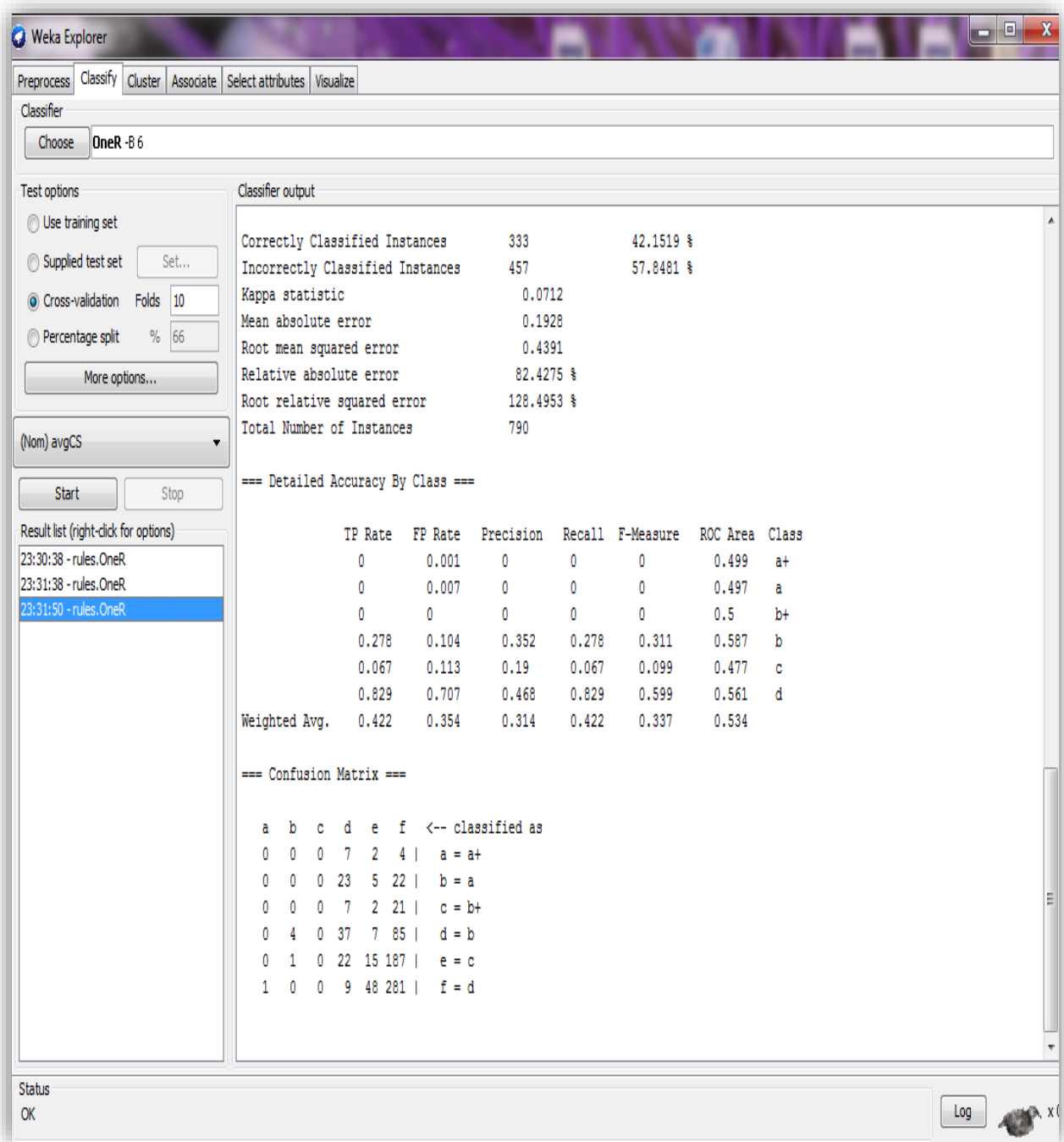


شكل (7.5) التصنيف

وتم استخدام إحدى خوارزميات التصنيف وهي (One-R Algorithm) كانت أكثر دقة وتأخذ وقت أقل والعوامل المستخدمة في الخوارزمية هي:

1-الكفاءة في توقع أداء الطالب .

2-القيم الإيجابية الصحيحة والقيم الإيجابية الخاطئة وقيم إل ROC



الشكل (8.5) النتائج المتحصل عليها بعد تطبيق خوارزمية التصنيف.

Correctly Classified Instances 333 42.1519 %

هذا السطر يوضح لنا مجموعة الحالات التي صنفها المصنف بشكل صحيح ونسبتها من مئة (instances)

Incorrectly Classified Instances 457 57.8481%

هذا السطر يوضح لنا مجموعة الحالات التي صنفها المصنف بشكل خطأ ونسبتها من مئة (instances)

لو نظرنا في أسفل النتائج لوجدنا مصفوفة تحوي على أرقام مختلفة

=== Confusion Matrix ===

a b c d e f <-- classified as

0 0 0 7 2 4 | a = a+

0 0 0 23 5 22 | b = a

0 0 0 7 2 21 | c = b+

0 4 0 35 7 15 | d = b

0 1 0 22 15 18 | e = c

1 0 0 9 48 28 | f = d

هذه المصفوفة تسمى مصفوفة الخطأ أو (Confusion Matrix) وهي مصفوفة مشهورة في مجال تعليم الآلة، تستخدم لاختبار أداء الخوارزميات. وهي عبارة عن جدول يحوي على معلومات وتفاصيل حول التصنيفات الفعلية (التي تم تصنيفها من قبل الإنسان) والتصنيفات التنبؤية التي تنبئها المصنف. فكل عامود في المصفوفة يمثل الفئات المتنبئة (Predicted class) و كل صف يمثل الفئات الفعلية (Actual class). ويتم تقييم أداء المصنف عادةً باستخدام البيانات الواردة في المصفوفة.

تبين الاعمدة عدد الطلبة و تبين الصفوف التقدير في مواد الحاسوب.

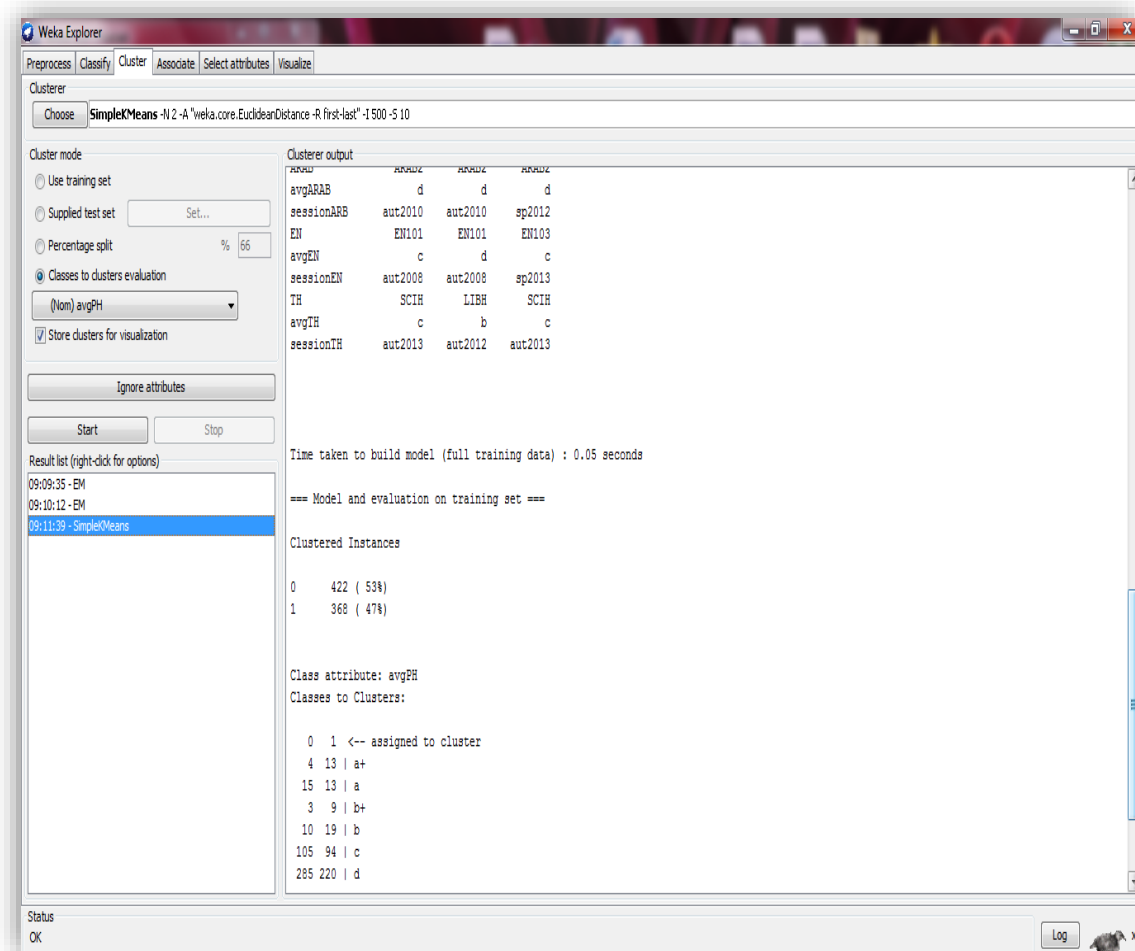
2.2.5 تقنية التجميع

هي احد فروع تنقيب البيانات تقسم خوارزمية [4] التجميع مجموعة البيانات أي عدة تجمعات حيث ان التشابه بين النقاط معين اكبر من التشابه بين النقطتين ضمن تجمعين مختلفين هناك العديد من الخوارزميات المستخدمة في عملية التجميع وابسط هذ الخوارزميات التي تم اختيارها

هي K Means خطوات هذه الخوارزمية هي

1. تحديد عدد التجمعات K، وهي تعتبر خطوة تهيئة اولية .

2. تحديد ادائيات مراكز التجمعات Centroid عشوائيا.



الشكل (9.5) النتائج المتحصل عليها بعد تطبيق خوارزمية التجميع

3.5 إحصائيات البيانات

قمنا بعمل مجموعة من المخططات بناءً علي البيانات المتحصل عليها وهي توضح الأداء في كل المواد على

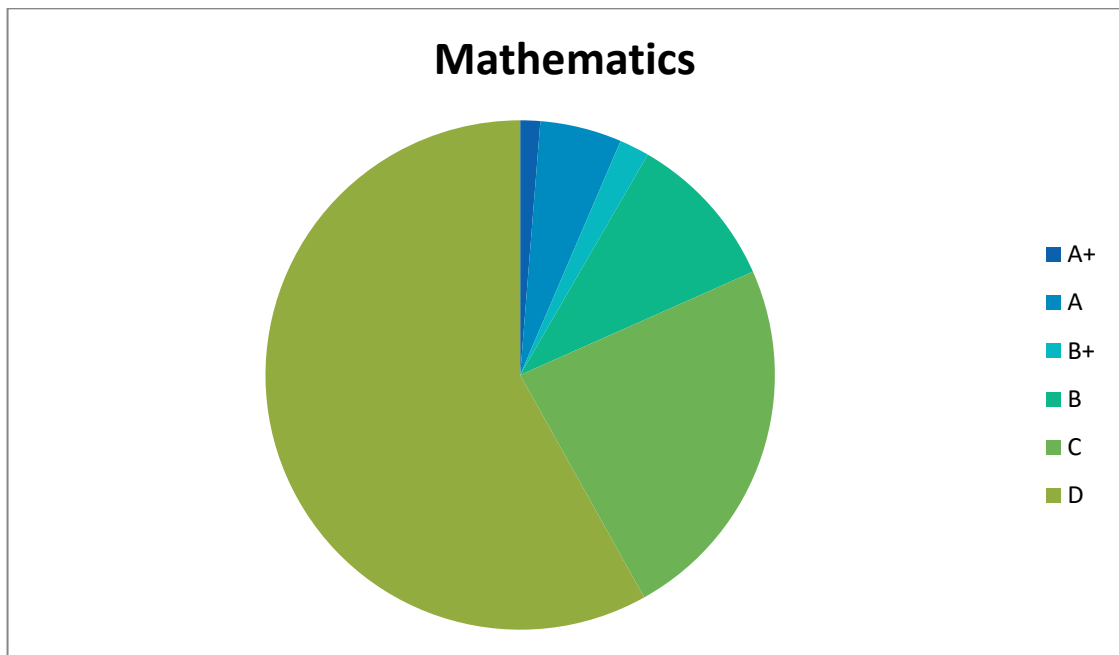
مدار الأربع سنين حيث

(A+,A) تمثل الممتاز

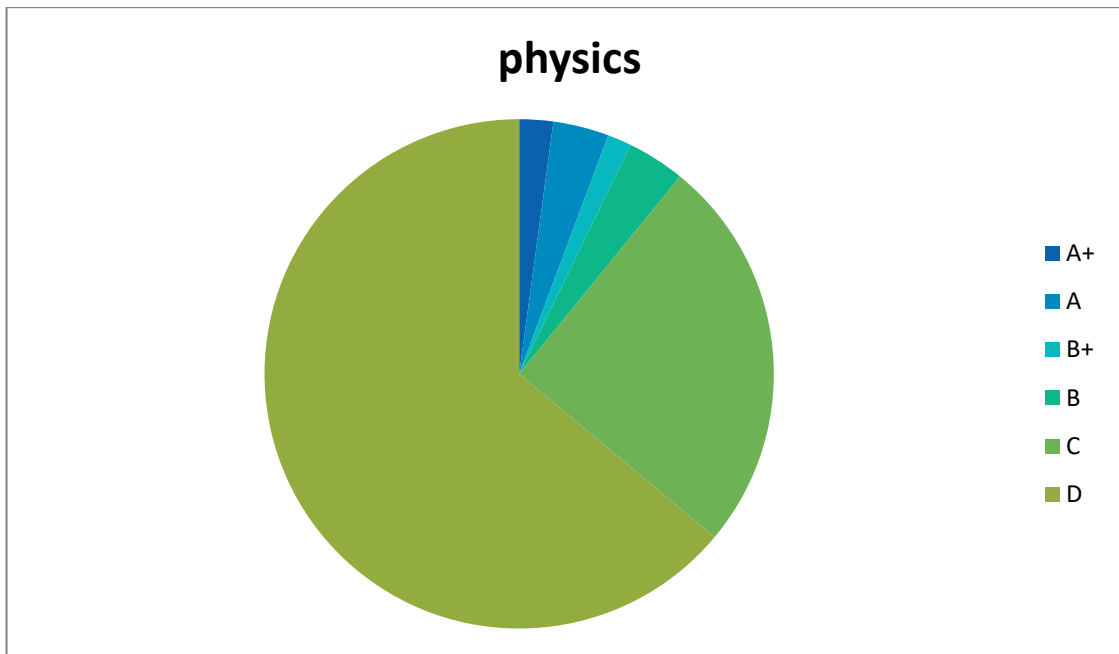
(B+,B) جيد جدا

(C) جيد

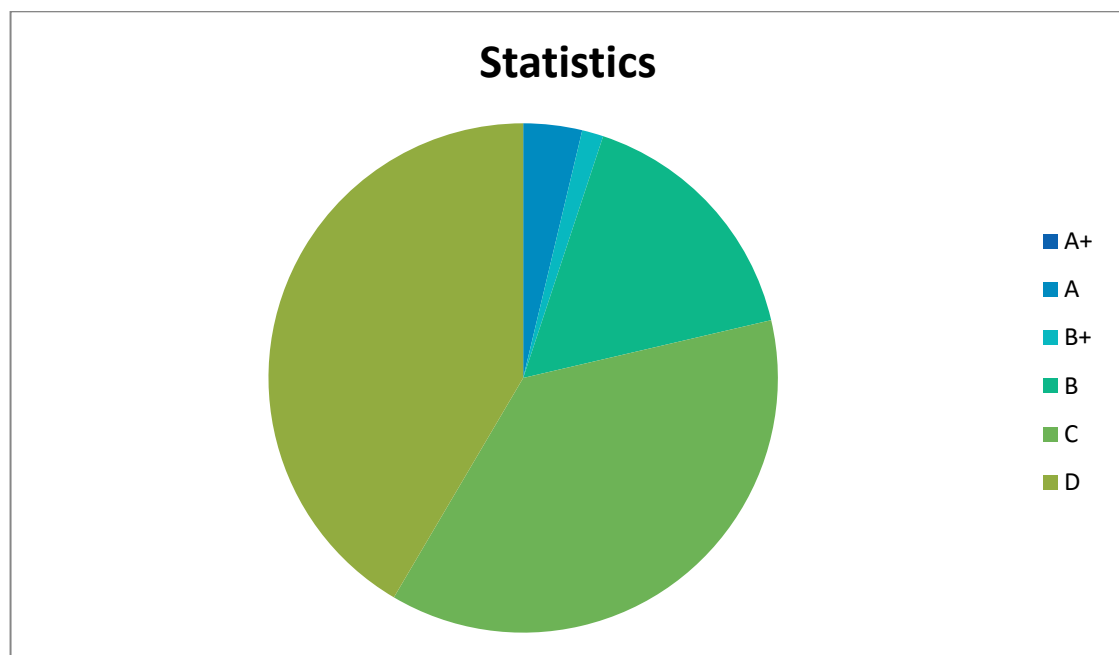
(D) مقبول



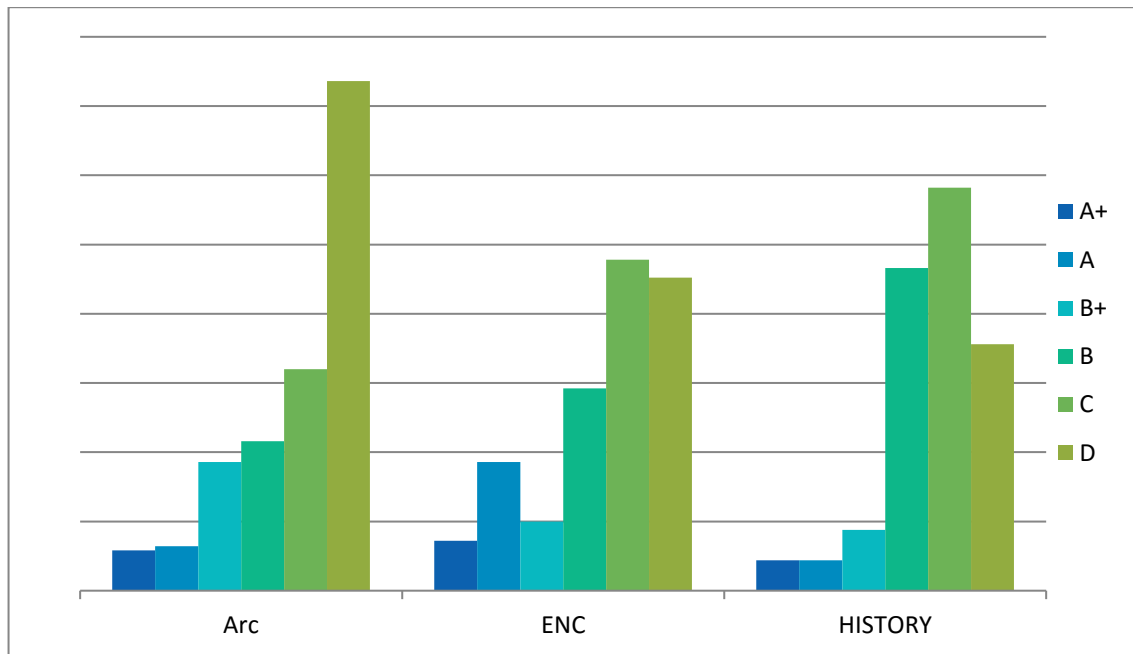
الشكل (10.5) يوضح أداء الطلبة مواد الرياضيات



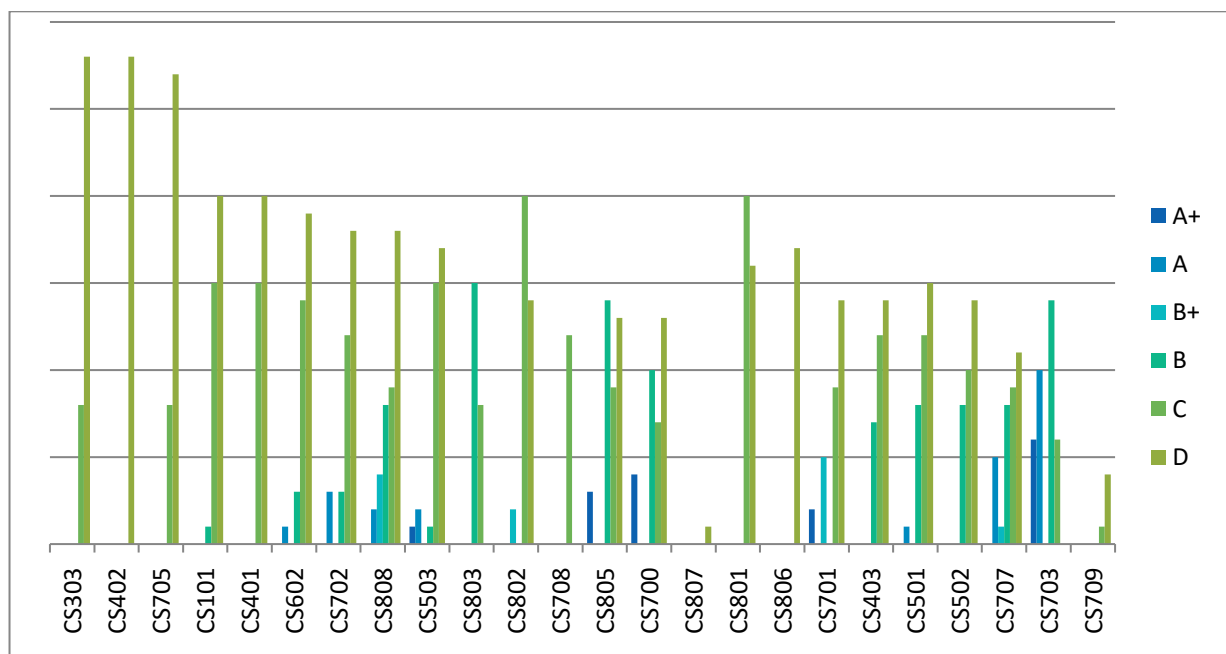
الشكل (11.5) يوضح أداء الطلبة في مواد الفيزياء



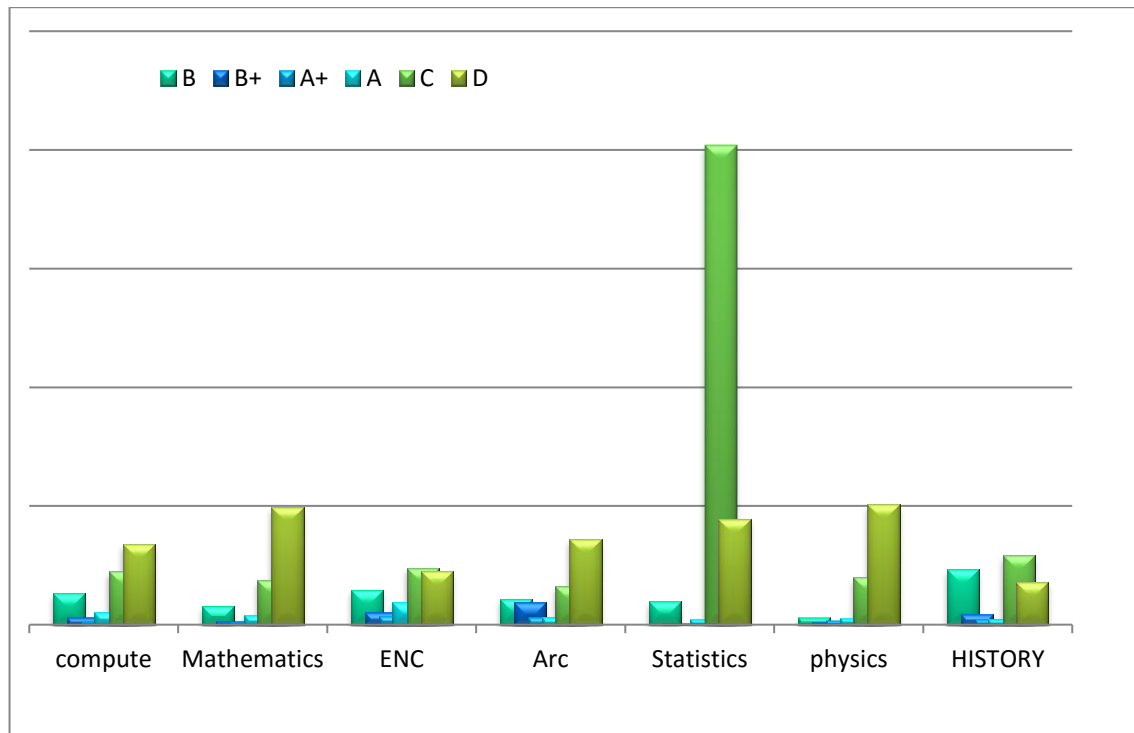
الشكل (12.5) يوضح أداء الطلبة في مواد الاحصاء



الشكل (13.5) يوضح أداء الطلبة في المواد الثقافية



الشكل (14.5) يوضح الأداء الطلبة في مواد الحاسوب



الشكل (15.5) يوضح اداء الطلبة خلال الاربع سنين

اتضح من خلال نتائج الخوارزميات المستخدمة في تقييم مستوى الطلاب من خلال درجاتهم اقل نسبة هي التقدير الممتاز في جميع المواد ماعدا المواد الثقافية واكبر نسبة هي التقدير المقبول مما بين الضعف في اداء الطالب

4.5التوصيات:

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

2 لزيادة كفاءة النتائج المتحصل عليها من هذا البحث نوصي بإضافة بيانات الطلاب الذين تعذر الحصول على بياناتهم

5.5 الخاتمة:

بعد الإطلاع على نتائج البحث وتحليلها ظهرت أهمية سجل الطالب الأكاديمي ومدى تأثيره علي المستويات الأكاديمية للطلاب قيد الدراسة ،وبالتالي ظهرت الحاجة إلي تحليل بيانات الطلاب للاستفادة منها في صنع القرارات ، حيث أن اجراءات 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 لموصول للأهداف المرجوة.

6.5 المراجع

- [1] التنقيب عن البيانات Data Mining . نبيل موصلي (2016) [1]
- [2] المرجع التعليمي في التنقيب عن البيانات . هالة الطويل, 2016 [2]
- بشير عباس، العلاق، الإدارة الرقمية: المجالات و التطبيقات، مركز الإمارات للدراسات و البحوث الإستراتيجية، ابوظبي، 2005، ص. 92 [3]
- [4] DATA MINING .THIRD EDITION (2006)
- [5]DATA MINING .THIRD EDITION(2012)
- [6]S. A. Kumar, “Mining Of Student Academic Evaluation Records in Higher Education,” pp. 67–70, 2012
- [7]V. Kumar and A. Chadha, “Mining Association Rules in Student ’ s Assessment Data,” vol. 9, no. 5, pp. 211–216, 2012
- [8]C. Romero, J. R. Romero, J. M. Luna, and S. Ventura, “Mining Rare Association Rules from ..e-Learning Data,” pp. 171–180, 2016
- [9]Introduction to Data Mining- by Tan, Steinbach, Kumar – 2004 .
- [10]Knitting and Crochet Patterns. (2015). Pattern Discovery In Data Mining
- Available at:
- <http://ktuliuepatt.com/pattern-discovery-in-data-mining/>

[11]Z.Kovacic "Early prediction of student success" Information Technology joint, Ca, Italy,
2010

[12] Bazsalica M., Naim P., Op. Cit., pp.68-69.

[13] Bazsalica M., Naim P., Op. Cit., pp.66.

[14] Bazsalica M., Naim P., Op. Cit., pp.68.