كلية علوم الحاسوب وتقانة المعلومات

نظام توصيه لمساعدة الإرشاد الأكاديمي

Recommendation System to help Academic Advising

مشروع مقدم كأحد متطلبات الحصول على
بكالوريوس الشرف في علوم الحاسوب

جامعة السودان للعلوم والتكنولوجيا
كلية علوم الحاسوب وتقانة المعلومات

نظام توصيه لمساعدة الإرشاد الأكاديمي

Recommendation System to help Academic Advising

مشروع مقدم كأحد متطلبات الحصول على
بكالوريوس الشرف في علوم الحاسوب

إعداد :

١. إيمان عبد الباسط مختار عوض الكريم

٢. سارة العاقب حسين أحمد

٣. عبير محمد عبد الماجد المكاوي

إشراف:
أ. وفاء فيصل مختار

أكتوبر ٢٠١٥

الآيـــــــــــة

بسم الله الرحمن الرحيم

(فَفَهَّمْنَاهَا سُلَيْمَانَ ۚ وَكُلًّا آتَيْنَا حُكْمًا وَعِلْمًا ۚ وَسَخَّرْنَا مَعَ دَاوُودَ الْجِبَالَ يُسَبِّحْنَ وَالطَّيْرَ ۚ وَكُنَّا فَاعِلِينَ)

صدق الله العظيم

سورة الأنبياء الآية (٧٩)

الحمد لله

الحمد لله الذي لا يحمد على مكروه سواه، الحمد لله الذي به تتم النعم وتكتمل الأعمال، الحمد لله عدد ما كان وعدد ما يكون وعدد الحركات والسكون، الحمد لله في الأولى والآخرة.

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

ربي أدم صلاتك الكاملة وبركاتك الشاملة وسلامك على سيد الخلق أجمعين محمد خاتم النبيين "صلى الله عليه وسلم" وعلى آلة وصحبه والتابعين، وتابعيهم ومن اهتدى بهدية إلى يوم الدين.

الإهـــــــــــــــــــــــــــداء

إلهي لا يطيب الليل إلا بشكرك ولا يطيب النهار إلى بطاعتك.. ولا تطيب اللحظات إلا بذكرك .. ولا تطيب الآخرة إلا بعفوك .. ولا تطيب الجنة إلا برؤيتك الله جل جلاله

إلى من بلغ الرسالة وأدى الأمانة .. ونصح الأمة .. إلى نبي الرحمة ونور العالمين

سيدنا محمد صلى الله عليه وسلم

إلى ملاكي في الحياة .. إلى معنى الحب وإلى معنى الحنان والتفاني .. إلى بسمة الحياة وسر الوجود
إلى من كان دعائها سر نجاحي وحنانها بلسم جراحي إلى أغلى الحبايب أمي الحبيبة

إلى من كلله الله بالهيبة والوقار .. إلى من علمني العطاء بدون انتظار .. إلى من أحمل أسمه بكل افتخار .. أرجو من الله أن يمد في عمرك لترى ثماراً قد حان قطافها بعد طول انتظار وستبقى كلماتك نجوم أهتدي بها اليوم وفي الغد وإلى الأبد

(والدي العزيز)

إلى أخواني ورفقاء دربي في هذه الحياة معكم أكون أنا وبدونكم أكون مثل لا شيء .. في نهاية مشواري أريد أن أشكركم على مواقفكم النبيلة إلى يا من تطلعت لنجاحي بنظرات الأمل

(أخوتي)

إلى الأخوات اللواتي لم تلدهن أمي .. إلى من تحلو بالإخاء وتميزوا بالوفاء والعطاء إلى ينابيع الصدق الصافي إلى من معهم سعدت ، وبرفقتهم في دروب الحياة الحلوة والحزينة سرت إلى من كانوا معي على طريق النجاح والخير إلى من عرفت كيف أجدهم وعلموني أن لا أضيعهم

(صديقاتي)

الشـــــكر والتقـــــدير

ولو أنني أوتيت كل بلاغة **** وأفنيت بحر النطق في النظم والنثر**
لما كنت بعد القول إلا مقصرا *** ومعترفا بالعجز عن واجب الشكر**

شكر مغلف بالامتنان و العرفان و أقول لك
أن للنجاحات أناس يقدرون معناه ، وللإبداع أناس يحصدونه ، لذا نقدّر جهودك
فلك منا كل الثناء والتقدير

الأستاذة وفاء فيصل مختار

إلي من أضاء بعلمه عقل غيره

واهدي بالجواب الصحيح حيرة سائلية

فاظهر بسماحته تواضع العلماء

وبرحابته سماحة العارفين

أستاذ أزهري فضل السيد

المستخلص

في هذا البحث تم استخدام بعض طرق نظم التوصية لتقديم المساعدة في عملية الإرشاد الأكاديمي بالنسبة للطالب الجامعي بكلية علوم الحاسوب وتقانة المعلومات بجامعة السودان للعلوم والتكنولوجيا. استخدامنا تقنيات جديدة لمعالجة وتحليل البيانات والتوصل لأفضل التوصيات وتقديمها للطلاب للحد من المشكلات، بما يقلل من عبء الإرشاد الأكاديمي على الأساتذة المرشدين. استخدمنا خوارزمية الجار الأقرب (K-NN) في التصفية التعاونية (collaborative filtering) من طرق نظم التوصية، وخوارزمية (Apriori) من قواعد الارتباط (association rule) في أداة تنقيب البيانات مفتوحة المصدر Weka، علي بيانات جمعت من استمارات التسجيل للطلبة المتعثرين في مسيرتهم التعليمية، وتم الحصول علي نتائجهم من نظام النتيجة. تحصلنا علي نتائج مفيدة لتجنب تكرار القرارات التي لم تفلح في حل مشاكل الطلاب فيما يخص خطتهم الدراسية أو القرارات التي أدت إلي مشاكل أخري للطلاب والتوصية بالقرارات التي نجحت في مساعدة الطلاب لتخطي مشاكلهم الدراسية والنجاح في مقررتهم والحصول علي نتائج أفضل. نسبة لقلة البيانات المستخدمة نحصل علي النتائج المتوقعة من خوارزمية Apriori وخوارزمية (K-NN)، وللحصول علي نتائج أفضل نوصي باستخدام كمية كبيرة من البيانات، بحيث يتوفر ملف كامل لكل طالب، ومن ثم استخدام البيانات في نظام خليط من خوارزميات التصفية التعاونية(collaborative filtering) وقواعد الارتباط(association rules).

Abstract:

In this research we proposed using some recommendation systems techniques that can help the Academic Advising process for the student of the Faculty of Computer Science and Information Technology at the University of Sudan for Science and Technology. New technologies were used to analyze the data and reach to the best recommendations and submit them to reduce the problems to meet the guidance needs of students and to reduce the burden of guidance councilors to teachers. We chose the nearest neighbor algorithm (KNN) from collaborative filtering recommendation systems techniques, and Aproiri algorithm of the association rules, applied in the data mining open source tool Weka. The data have been collected from the registration forms of students who have difficulties, their results were obtained from the records of the results system. Some recommendations results were obtained on directing and guiding the students that might help to avoid the repeating of inappropriate decisions that have not succeeded in solving the problems of students with regard to their plan of study or decisions that led to other problems for students, and some recommendation to help overcome their problems and academic success and getting the best results. Due to small size of information collected, we have not got the expected results after applying the two algorithms, we recommend using of a hybrid technique between KNN and Apriori and be applied on large database consisting of a complete record for each student.

المصطلحات الموحدة

المصطلح	تعريفه
الساعات المعتمدة	عبارة عن وحده تعليمية يقصد بها ساعة محاضره أو ساعتان إلى ثلاث من ساعات المتابعة أو العملي في الأسبوع لمدة خمسة عشر أسبوع (فصل دراسي).
السجل الأكاديمي	هو كشف يبين المقررات التي درسها الطالب في تسلسل والتقديرات التي أحرزها لهذه المقررات والمعدل الفصلي والتراكمي والمعدل التراكمي في كل فصل دراسي.
الدرجة النقطية للمقرر:	يقصد بها درجة الطالب في المقرر من100 مقسومة على 25 (أو على 30لإزالة الرسوب) مقربه لمنزله عشريه واحده.
نقاط المقرر	تساوي حاصل ضرب الدرجة النقطية للمقرر في عدد الساعات المعتمدة.
الساعات المعتمدة الفصلية للطالب	هي مجموع الساعات المعتمدة للمقررات التي درسه الطالب في الفصل الدراسي المعين وجلس للامتحانات فيه مضافا إليها ساعات المقررات التي جلس لها امتحانا بديلا أو لإزالة الرسوب مع امتحانات الفصل
النقاط الفصلية للطالب :	هي مجموع نقاط المقررات التي درسها الطالب في الفصل الدراسي المعين وجلس للامتحانات فيه مضافا إليها نقاط المقررات التي جاس لها امتحانا بديلا أو لإزالة الرسوب مع امتحانات الفصل
المعدل الفصلي للطالب:	يقصد به حاصل قسمة النقاط الفصلية للطالب على الساعات المعتمدة الفصلية للطالب على الساعات المعتمدة الفصلية للطالب مقربه لمنزلتين عشريتين
الساعات المعتمدة التراكمية للطالب :	هي مجموع الساعات المعتمدة التراكمية للمقرات التي درسها الطالب وجلس للامتحان فيها لفصلين دراسيين على الأقل.
النقاط التراكمية للطالب	هي مجموع نقاط المقررات التي سبق للطالب دراستها والجلوس للامتحان فيها لفصلين دراسيين على الاقل مضافا إليها أي مقرر يكرر الطالب الجلوس للامتحان فيه بعدد مرات التكرار .
المعدل التراكمي للطالب	يقصد به حاصل قسمة النقاط التراكمية على الساعات التراكمية للطالب مقربه لمنزلتين عشريتين.
الإنذار الأكاديمي	هو ان يخضع الطالب للإنذار بالفصل من الدر اسه إذا قل معدله التراكمي عن 2.00.

فـهـرس الأشكـال

رقم الشكل	موضوع الشكل	رقم الصفحه
١.٢	خطوات اكتشاف المعرفة من قواعد البيانات	٤
٢.٢	واجهة Rattle	٢٢
٣.٢	واجهة (Rapid Miner)	٢٢
٤.٢	واجهة Apatch Mahout	٢٣
٥.٢	واجهة pspp	٢٤
٦.٢	واجهة R	٢٥
٧.٢	شاشة بدء تشغيلWEKA	٢٦
٨.٢	شاشة المستكشف الرسومية قبل الاختيار	٢٦
٩.٢	المستكشف الرسوميه بعد الاختيار	٢٧
١٠.٢	شاشة (classify)	٢٤
١.٥	استمارة الطالب الأكاديمي	٤١
٢.٥	نموذج لبيانات (apriori)	٤٢
٣.٥	نموذج لبيانات (knn)	٤٢
٤.٥	طريقة إختيار خوارزمية(Apriori)	٤٣
٥.٥	شاشة إختيار(support) و(confidence)	٤٤
٦.٥	شاشة (classify)	٤٨
١.٦	نتائج الخوارزمية(Apriori)	٤٨
٢.٦	مخطط نسب خوارزمية(KNN)	٤٩

فهرس المحتويات

الباب الأول المقدمة	الصفحة
١.١مقدمة.	١
٢.١ مشكلة وأهمية البحث.	١
١.٣أهداف البحث.	١
١.٤منهجية البحث.	١
٥.١ حدود البحث.	٢
٦.١ هيكلية البحث.	٢
الباب الثاني تنقيب البيانات	
الفصل الأول تنقيب البيانات	
٢.١.٢مقدمه	٦
٢.١.٢خطوات استكشاف المعرفة من قواعد البيانات	٦
٢.١.١.٢اكتشاف البيانات	٦
٢.٢.١.٢تصفية البيانات	٦
٣.٢.١.٢ تكامل البيانات	٧
٤.٢.١.٢ اختيار البيانات	٧
٥.٢.١.٢ تحويل البيانات	٧
٦.٢.١.٢التنقيب عن البيانات	٧
٧.٢.١.٢ تقييم النمط	٨
٨.٢.١.٢ تمثيل المعرفة وتقديمها	٨
٣.١.٢ أهداف التنقيب	٨
٤.١.٢ أنواع تنقيب البيانات	٨
١.٤.١.٢ التنقيب التنبؤي	٨
٢.٤.١.٢التنقيب الوصفي	٨
٥.١.٢مهام تنقيب البيانات	٨
١.٥.١.٢ العنقدة	٨
٢.٥.١.٢اكتشاف التسلسل	٩
٣.٥.١.٢ التصنيف	٩
٤.٥.١.٢ الانحدار	٩
٥.٥.١.٢التخليص	٩
٦.٥.١.٢ قاعدة الارتباط	٩
١.٦.٥.١.٢قياسات قواعد الارتباط	١١
٢.٦.٥.١.٢ خوارزمية (Apriori)	١٢
٦.١.٢ وسائل التنقيب عن البيانات	١٢
١.٦.١.٢الاستدلال المبني علي الحالات	١٢
١.١.٦.١.٢ الكشف عن القانون(Rule Discovery)	١٢
٢.١.٦.١.٢معالجة الإشارات (Signal Processing)	١٢
٣.١.٦.١.٢ شبكات عصبيه (Neural Net)	١٢
٤.١.٦.١.٢ منحنيات غير ثابتة (Fractals)	١٢
الفصل الثاني نظم التوصية	
١.٢.٢مقدمه	١٤
٢.٢.٢ تاريخ نظم التوصية	١٤

٣.٢.٢ نظم التوصية	١٥
٤.٢.٢ أنواع نظم التوصية	١٥
١.٤.٢.٢ الفلترة التعاونية (Collaborative Filtering)	١٦
١.١.٤.٢.٢ أنواع الفلترة التعاونية	١٧
٢.١.٤.٢.٢ تحديات الفلترة التعاونية	١٧
٢.٤.٢.٢ الفلترة على المحتوى (Content-based filtering)	١٨
٣.٤.٢.٢ نظم التوصية المختلطة (Hybrid Recommender System)	١٨
الفصل الثالث أدوات تنقيب البيانات	
١.٣.٢ مقدمه	٢٠
٢.٣.٢ Rattle	٢٠
٣.٣.٢ Rapid Miner	٢١
٤.٣.٢ Apache Mahout	٢١
٥.٣.٢PSPP	٢١
٦.٣.٢R	٢٢
٧.٣.٢ Weka	٢٣
١.٧.٣.٢ ملف (ARFF)	٢٤
٢.٧.٣.٢ تحميل البيانات في weka	٢٤
الباب الثالث الدراسات السابقة	
١.٣دراسة أولى	٢٧
٢.٣دراسة ثانيه	٢٩
٣.٣ دراسة ثالثه	٣٠
٤.٣ دراسة رابعة	٣١
الباب الرابع الإرشاد الأكاديمي	
١.٤مقدمه	٣٣
٢.٤ الإرشاد الأكاديمي	٣٣
٣.٤ أهمية الإرشاد الأكاديمي	٣٣
٤.٤المرشد الأكاديمي	٣٣
٥.٤مهام المرشد الأكاديمي	٣٤
٦.٤ المشاكل التي تواجه المرشد الأكاديمي	٣٤
٧.٤ بعض السمات الأساسية للساعات المعتمدة	٣٤
٨.٤ حساب النتيجة	٣٥
الباب الخامس جمع البيانات	
١.٥ مقدمة	٣٧
٢.٥ اختيار البيانات	٣٨
٣.٥ تهيئة البيانات لقواعد الارتباط	٣٨
٤.٥ تهيئة البيانات للفلترة التعاونية	٣٩
٥.٥تنقيب البيانات	٣٩
٦.٥ طريقة إختيار خوارزمية ال (Apriori)	٤٠
٧.٥ كيفية عمل خوارزمية (K-NN)	٤٢
الباب السادس النتائج والتوصيات	
١.٦ مقدمة	٤٥
٢.٦ نتائج خوارزمية(Apriori)	٤٥
١.٢.٦مناقشة نتائج خوارزمية Apriori	٤٥
٢.٢.٦ التوصيات	٤٦

٣.٦مخطط يوضح نسب خوارزمية (KNN)	٤٦
١.٣.٦ مناقشة نتائج المخطط	٤٧
٢.٣.٦التوصيات	٤٧
الخاتمة	٤٨
المراجع	٥٠

الباب الأول

المقدمة

١.١ المقدمة

مع زيادة أهمية الإرشاد الأكاديمي لنجاح الطالب ونجاح نظام الساعات المعتمدة الذي يعتمد نجاحه بالكامل علي الإرشاد الجيد لتعريف الطلاب بالأنظمة الدراسية، ومساعدتهم على اتخاذ القرارات التي تتناسب مع قدراتهم وإمكاناتهم، وكذلك معاونتهم على السير في الدراسة على أفضل وجه ممكن، والتغلب على ما يعترضهم من عقبات، ولتمكين الطالب على التكيف الأكاديمي والنفسي مع الواقع التعليمي لهذا النظام الذي يختلف كليًا عما اعتاده الطالب في المراحل السابقة لدراسته الجامعية، أصبح من الضروري إيجاد تقنيات وطرق ووسائل تساعد المرشد الأكاديمي في القيام بمهمته لتقديم التوصيات باستخدام تطبيقات الحاسوب،وجاءت فكرة نظم الـتوصية(recommender system) بطرق ذكية للمساعدة في حل المشاكل واتخاذ القرارات.ومن أجل تقديم التوصيات المناسبة، يجب على النظام أن يتعلم أولاً تفضيلات المستخدم من خلال تحليل الأنشطة التصنيفية أو التوصيات الماضية.

١.٢مشكلة البحث

عدم توفر التدريب الكافي للمرشدين الأكاديميين والاعتماد على قدراتهم الشخصية في إكتساب الخبرة في الإرشاد كما يمثل الإرشاد عبئاً إدارياً إضافيا. ويشكو بعض المرشدين من زيادة عدد طلابهم مما يتعذر على المرشد الأكاديمي أحياناً القيام بمهمة الإرشاد بسبب كثرة المهام المسندة إليه أو لأي سبب آخر.

١.٣أهداف البحث

استخدام تقنيات جديدة لمساعدة المرشد الأكاديمي بتحليل البيانات والتوصل لأفضل التوصيات وتقديمها للحد من المشكلات بما يلبي الحاجات الإرشادية للطالب والتقليل من عبء الإرشاد على الأساتذة المرشدين، وتوجيه الطالب ومساعدته في اختيار المقررات الدراسية المناسبة في كل فصل.

١.٤منهجية البحث

في هذا المشروع سيتم إقتراح طريقة جديدة لمساعدة المرشد الأكاديمي قائمة علي الطرق المستخدمة حالياً في نظم التوصية (recommender systems). هذه الطريقة المقترحة تهدف إلي بناء قرارات وتنبؤات ذات صلة بالارشاد الأكاديمي.

١.٥ حدود البحث

تشمل الدراسة تطبيق طريقه واحده من طرق نظم التوصية(collaborative filtering) على الإشراف الأكاديمي وذلك بإستخدام خوارزمية الجار الأقرب(K-NN)، وخوارزمية قواعد الإرتباط في تنقيب البيانات وهي (apriori).

تتركز الدراسة على جامعة السودان للعلوم والتكنولوجيا في جوانب الإرشاد الأكاديمي، وقد إستخدمنا إستمارات طلاب كلية علوم الحاسوب وتقانة المعلومات في السنوات السابقة.

وقد تم إجراء هذا البحث في العام الدراسي ٢٠١٥م، بإجراء مقابلات شخصية مع مسئولين على الإشراف الأكاديمي والإداري، والشئون العلمية.

١.٦ هيكلية البحث

يتضمن البحث بالإضافة إلى هذا الباب الأبواب :-

الباب الثاني : ويتضمن ثلاثة فصول تنقيب البيانات ونظم التوصية وأدوات تنقيب البيانات.

الباب الثالث : الدراسات السابقة

الباب الرابع: الإرشاد الأكاديمي.

الباب الخامس: منهجية البحث.

الباب السادس : مناقشة النتائج والتوصيات.

الباب الثاني

تنقيب البيانات

نظم التوصية

أدوات تنقيب البيانات

٢.١ الفصل الأول: تنقيب البيانات

٢.٢ الفصل الثاني : نظم التوصية

٢.٣ الفصل الثالث: أدوات التنقيب البيانات

الفصل الأول

تنقيب البيانات

١.١.٢ مقدمة

التنقيب في البيانات هو عملية تحليل كمية بيانات (عادة ما تكون كمية كبيرة) لإيجاد علاقة منطقية تلخص البيانات بطريقة جديدة تكون مفهومة ومفيدة لصاحب البيانات. يطلق اسم "نماذج models " على العلاقات والبيانات الملخصة التي يتم الحصول عليها من التنقيب في البيانات. يتعامل تنقيب البيانات عادة مع بيانات قد تم الحصول عليها بغرض غير غرض التنقيب في البيانات (مثلاً قاعدة البيانات لتعاملات في مصرف ما) مما يعني أن طريقة التنقيب في البيانات لا تؤثر مطلقاً على طريقة تجميع البيانات ذاتها. هذه هي أحد النواحي التي يختلف فيها التنقيب في البيانات عن الإحصاء. يشير التعريف أيضاً إلى أن كمية البيانات تكون عادة كبيرة، أما في حال كون كمية البيانات صغيرة فيفضل استخدام الطرق الإحصائية العادية في تحليلها[1].

٢.١.٢ خطوات استكشاف المعرفة من قواعد البيانات

Knowledge Discovery in Database (KDD):

يعد تنقيب البيانات (Data Mining) مرحلة من مراحل اكتشاف المعرفة في قواعد البيانات[2]**التي تشير الى استخراج المفاهيم الضمنية غير الاعتيادية والتي لم تكن معروفة سابقا** .اكتشاف المعرفة في قواعد البيانات (KDD) ليس بالعملية السهلة والتي قد يعتقد البعض أنها تتوقف عند تجميع البيانات وإدارتها، بل نراها تمتد إلى التحليل والتوقع والتنبؤ بما سيحدث مستقبلاً.**وتتضمن خطوات (KDD) عدد من المراحل تبدأ من جمع البيانات الخام الى مرحلة الحصول على المعرفة الجديدة كما موضح بالشكل (١.٢):**

١.٢.١.٢ اكتشاف البيانات (Data discovery):

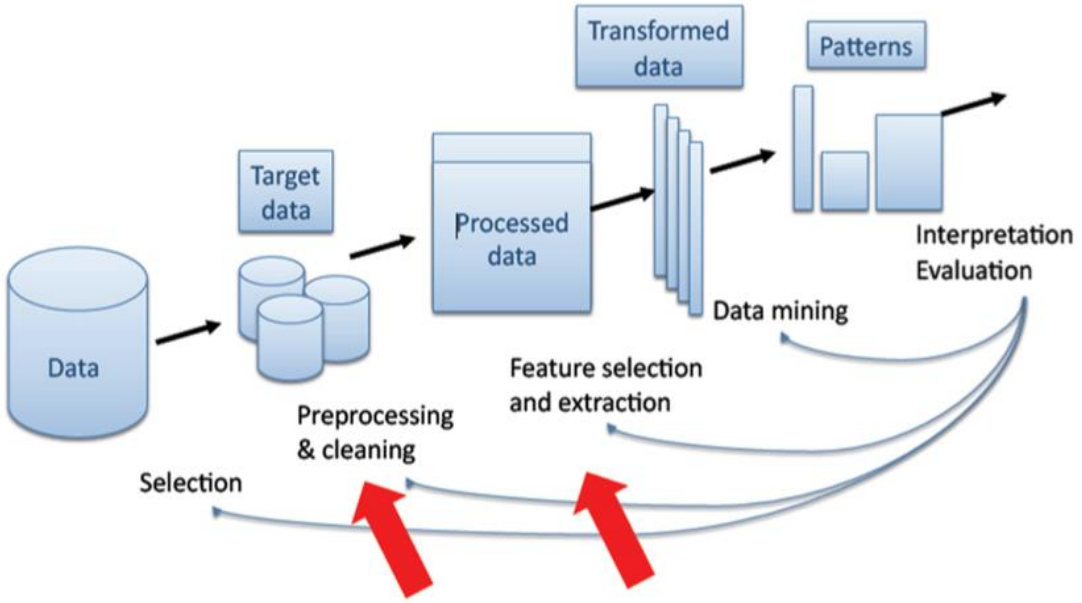
وهي مرحلة جمع البيانات وتشمل كشف وتحديد وتوصيف البيانات المتاحة.

٢.٢.١.٢ تصفية البيانات (Data filtering):

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

٣.٢.١.٢ تكامل البيانات(DataIntegration):

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



شكل(1.2)خطوات اكتشاف المعرفة من قواعد البيانات

٤.٢.١.٢ اختيار البيانات(Data selection):

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

٥.٢.١.٢ تحويل البيانات(Data Conversion):

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

٦.٢.١.٢ التنقيب عن البيانات(Data mining):

أي استخدام طرق ذكية تطبق لاستخلاص أنماط البيانات لاستخراج نماذج مفيدة قدر الإمكان.

٧.٢.١.٢ تقييم النمط (Evaluate pattern) :

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

٨.٢.١.٢ تمثيل المعرفة وتقديمها

(Knowledge representation and submission):

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.

٣.١.٢أهداف التنقيب

هناك ثلاثة أهداف للتنقيب في البيانات:[1]

١ - من أجل تعليل بعض الظواهر المرئية.

٢- من أجل التثبت من نظرية ما. مثال: التثبت من النظرية التي تقول بأن الأسر الكبيرة تهتم بالضمان الصحي أكثر من الأسر الصغيرة عددا.

٣- من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة .مثال: كيف سيكون الانفاق العام إن كان ملازما لعمليات خداع واسعة من قبل البطاقات الائتمانية

٤.٢ أنواع التنقيب في البيانات

يوجد نوعين من تنقيب البيانات وهما التنقيب التنبؤي والتنقيب الوصفي.

١.٤.١.٢ التنقيب التنبؤي

يحاول إيجاد أفضل التنبؤات اعتمادا 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.

٢.٤.١.٢ التنقيب الوصفي

ينتج عنه معلومات جديدة بناء على المعلومات الموجودة داخل البيانات المستخدمة في عملية التنقيب.

٥.١.٢ مهام تنقيب البيانات

يوجد العديد من مهام تنقيب البيانات ومنها:

١.٥.١.٢ 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 اعتمادا على اشتراكها بالخواص المتشابهة وأن العنقدة هي تقسيم غير موجه للبيانات ، وهي عكس التصنيف ، كما أنها تساعد المستفيد على فهم التركيب الطبيعي للمجموعات من البيانات.

٢.٥.١.٢ اكتشاف التسلسل

في هذه الطريقة يتم البحث لاكتشاف نماذج تحدث بالتسلسل اذ تكون المدخلات عبارة عن بيانات تشكل مجموعة متسلسلة و تستخدم تقنيات التسلسل التحليل القائم على الوقت لانتزاع 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.

٣.٥.١.٢ التصنيف

يستخدم التصنيف بشكل واسع في حل كثير من المشكلات من خلال تحليل مجموعه من البيانات ووضعها في شكل أصناف أو أقسام يمكن استخدامها فيما بعد لتصنيف البيانات المستقبلية وهناك عدد من الطرق التي يمكن استخدامها في تصنيف البيانات باستخدام الخوارزميات مثل الخوارزمية الإحصائية، وخوارزميات الشبكات العصبية، والعديد من خوارزميات تعلم الالة.

٤.٥.١.٢ الإنحدار

يستخدم للعثور على دالة تنمذج البيانات بأقل خطأ.

٥.٥.١.٢ التلخيص

يوفر عرضا محكما للبيانات يحتوي علي التصور الكامل وتوليد التقرير

٦.٥.١.٢ قاعدة الارتباط

هي أحد التقنيات الواعدة من تنقيب البيانات كاداه 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، بمعني آخر أنها قواعد ارتباطيه معينه بين مجموعه من البيانات في قواعد البيانات.

١.٦.٥.١.٢ قياسات قواعد الارتباط

تعتبر قاعدة الارتباط (Association Rule) [3]أحد أهم مجالات التنقيب في البيانات فهي طريقة بحث جيدة وعامة لاكتشاف العلاقات الهامة بين المتحولات في قواعد البيانات الضخمة بإنتاج قواعد ارتباط لاكتشاف الانتظام بين المنتجات في بيانات العمليات الكبيرة فإن قواعد الارتباط تستخدم في تطبيقات عديدة مثل التنقيب في الـويب ومجال الـمعلومات الحيوية (bioinformatics).

تعريف أساسي لمسألة التنقيب عن قاعدة الارتباط حيث لتكن لدينا المجموعة I={i1,i2,i3,i4,…in} (items) تدعى خاصية المواد(n)

ولتكن لدينا مجموعة من العمليات (transactions) : والتي هي بمثابة قاعدة البيانات

$$D=\{t1,t2,,,\ldots tm\}$$

حيث كل عملية (ti) من المجموعة (D) لها اسم يميزها وتحوي مجموعة جزئية من المواد الموجودة في المجموعة (I).

عندئذ القاعدة (Rule) تعرف بالصيغة:

$$x \Rightarrow y \text{ where } x , y \in I \text{ and } x \cap y = \emptyset$$

لتوضيح هذه المفاهيم نستخدم مثال صغير من السوق المركزي (supermarket) حيث مجموعة المواد هيI{ milk , bread , butter ,beer, cheese } = :

وقاعدة بيانات صغيرة تتضمن المواد (items) حيث:

[الرمز (١) موجود ، الرمز (٠) غائب عن الصفات (المادة غير موجودة)]

مثلاً القاعدة في هذه الحالة هي{ butter } ⇒ { milk , bread } :

ذلك يعني أنه إذا تم شراء الحليب والخبز فإن الزبون أيضاً يشتري الزبدة، لذلك فإن مجموعات المواد (items) تعتبر هامة من الناحية الإحصائية ومجموعات البيانات (datasets)غالباً تحوي آلاف أو ملايين من مجموعات المواد ولاختيار القواعد الهامة من مجموعة القواعد المحتملة يوجد بعض القيود الهامة أفضل هذه القيود هي:

-الدعم (support)ومعامل الثقة (confidence).

- الدعم (support)

الدعم(supp(x لمجموعة المواد (x) يدل على عدد مرات تكرار القاعدة في جميع المعاملات في قاعدة البيانات.ففي المثال لدينا : مجموعة المواد {milk , bread } = item set .

supp(x)=(xUy)/n

حيث**n** =المجموع الكلي للمعاملات.

SUPPORT =2/4=0.5

أي يمكن أن تحدث بمقدار 50%

لكل عملية (transactions).

- معامل الثقة (confidence)

معامل الثقة للقاعدة (confidence) يعرف بالعلاقة:

Conf{ x⇒ y}= SUPP(X ∪ Y)/ SUPP(X)

ففي مثالنا السابق مدى: confidence {milk , bread } ⇒{butte}

عند ظهور { milk , bread , butter} مره واحدة في قاعدة البيانات

Confidence=1/2=0.5

ومن خوارزميات قواعد الارتباط خوارزمية **(Apriori)**وفيما يلي سنتحدث عنها بالتفصيل:

٢.٦.٥.١.٢ خوارزمية (Apriori)

هي خوارزمية للتنقيب عن العناصر المتكررة وتطبيق قوانين الارتباط على (transactional databases) وتعمل عن طريق تحديد العناصر الفردية المتكررة في قاعده البيانات وتوسيعها لتشمل مجموعات اكبر. واكبر عنصر هو العنصر الاكثر ظهورا في قاعدة البيانات ومجموعات العنصر المتكررة يتم تحديدها عن طريق خوارزمية (Aproiri)[4] التي تستخدم لتحديد قوانين الارتباط، هذه الخوارزمية مصممة للعمل علي قواعد البيانات التي تحتوي علي (transactions) وتم تصميم الخوارزميات الاخري لايجاد قوانين الارتباط في حالة عدم وجود (transactions) 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 اللحظة المعينة وهي خطوة تولد الفلترة (candidate generation) ويتم إختبار هذه المجموعات المرشحة. وتنتهي هذة الخوارزمية عندما لا يتم العثور علي اي تمديدات اخري ناجحه.

$\text{Apriori}(T, \epsilon)$
$L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}$
$k \leftarrow 2$
while $L_{k-1} \neq \emptyset$
$C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k-1\} \not\subseteq L_{k-1}\}$
for transactions $t \in T$
$C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$
for candidates $c \in C_t$
$count[c] \leftarrow count[c] + 1$
$L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}$
$k \leftarrow k + 1$
return $\bigcup_k L_k$

توضيح رموز خوارزمية (Apriori):

١. (T)=Transaction.
٢. (€)=support.
٣. (Cκ)= المجموعة المرشحة في المستوىk .
٤. ([Count[c])= حقل هيكلة البيانات التي تمثل المجموعات المرشحة c ،الذي من المفترض أن تكون قيمتة الإبتدائية تساوي صفرا.
٥. (L) = اكبر بند(Large item set).

تعمل خوازمية (Apriori) بإيجاد مجموعات العناصر المتكرره (العناصر التي ظهرت في قاعدة البيانات والتي أكبر من أو يساوي الحد الأدنى من حد الدعم)،ونولد مرشح من عناصر متكررة، ومن ثم تهذيب النتائج للعثور على مجموعات العنصر المتكررة. وتوليد قواعد ارتباط قوية من مجموعات العنصر المتكررة

٦.١.٢ وسائل التنقيب في البيانات

هناك عدة وسائل مختلفة من أجل التنقيب في البيانات. اختيار الوسيلة المناسبة يعتمد على طبيعة البيانات تحت الدراسة وعلى حجمها. يمكن اجراء عملية التنقيب في البيانات بالمقارنة مع سوق البيانات ومخزن البيانات.بعض من هذه الوسائل هي: (كل وسيلة يتبعها وظيفة).

١.٦.١.٢ الاستدلال المبني على الحالات

الاستدلال المبني على الحالات(Case-Based reasoning) التفكير واستخلاص النتائج والقوانين من أمثلة حية و قضايا تم حلها سابقا.

١.١.٦.١.٢ الكشف عن قانون (Rule Discovery)

البحث عن منوال معين أو علاقة معينة في جزئية كبيرة من البيانات.

٢.١.٦.١.٢ معالجة الإشارات (Signal Processing)

ايجاد الظواهر المتشابه مع بعضها البعض.

٣.١.٦.١.٢ شبكات عصبيه(Neural Net)

وهي نماذج قابلة لتنبؤ النتائج، تم تطويرها على أسس الإستنباط من عقل الإنسان.

٤.١.٦.١.٢ منحنيات غير ثابتة(Fractals)

تصغير البيانات الكبيرة من دون ضياع المعلومات.

الفصل الثانـي

نظم التوصية

١.٢.٢مقدمة

قدرة أجهزة الكمبيوتر لتقديم التوصيات في تاريخ الحوسبة كانت في وقت مبكر، هذه الحواسيب تستند علي(librarian)، وهي خطوة مبكرة نحو نظام المذكي الآلي(automatic system recommender). وكان بدائي إلي حد ما، ولكنه قدم إدخال مبكر هام في فضاء التوصيات.

٢.٢.٢تاريخ نظم التوصية (Recommendation System)

في أوائل١٩٩٠ بدأت الفلترة التعاونية (collaborative filtering) [5]للتعامل مع المعلومات الزائدة في أماكن المعلومات علي الإنترنت.كان نظام الفلترة التعاونية اليدوي يسمح للمستخدم بالإستعلام عن العناصر الموجودة في مجال المعلومات، مثل البريد الإلكتروني للشركات،على أساس آراء المستخدمين الآخرين أوالإجراءات وهي لا تتطلب جهداً من جانب مستخدميها، وأيضاً سمح لهم بإلاستفادة من ردود فعل القراء السابقة لتحديد أهميتها بالنسبة لهم.

وسرعان ما تبع نظم الفلترة التعاونية الآلية (Automated collaborative filtering systems) تحديد الآراء ذات الصلة أوتوماتيكياً وتجميعها لتقديم التوصيات وتستخدم هذه التقنية لتحدد المواد من الأعضاء التي من المحتمل أن تكون مثيرة للإهتمام لمستخدم معين، يحتاج المستخدمين فقط إلى توفير تقييمات أو تنفيذ إجراءات أخرى يمكن ملاحظتها; النظام يجمع بين هذه الآراء مع هذا التقييم أو تصرفات المستخدمين الآخرين لتقديم نتائج ذات طابع شخصي.مع هذه الأنظمة لا يحتاج المستخدمون الحصول على أي معرفة مباشرة لآراء المستخدمين الآخرين، كما أنها لا تحتاج إلى معرفة مستخدمين أوغيرهم من البنود في النظام من أجل الحصول على التوصيات.

خلال هذا الوقت نظم التوصية والفلترة المتعاونة أصبحا موضوعاً مثيراً للإهتمام المتزايد بين تفاعل الإنسان والحاسوب، والتعلم الآلي، وإسترجاع المعلومات للباحثين، وأنتج هذا الإهتمام عدداً من نظم التوصية لمختلف المجالات، مثل (Ringo) للموسيقى، (Core Video Recommender the Bell) للأفلام، والأدب و التسويق وتحليل التوصية للقدرة على زيادة المبيعات.

في أواخر ١٩٩٠ التطورات التجارية من تقنيات التوصية بدأت في الظهور.ربما كان التطبيق الأكثر شهرة على نطاق واسع من تقنيات نظم التوصية هي الأمازون(Amazon.Com)، وهذا البند للمستخدم هو يشاهد حالياً، فإنه يوصي البنود للمستخدم للنظرة الشرائية.

نظم التوصية غيرت طريقة الناس في العثور على المنتجات والمعلومات. بدراسة أنماط السلوك لمعرفة ما يفضله شخص ما من بين مجموعة من الأشياء التي لم يجربها. تطورت التقنية الكامنة وراء أنظمة التوصية على مدى السنوات العشرين إلى مجموعة غنية من الأدوات التي تمكن الباحث لتطوير توصية فعالة .

٣.٢.٢ نظم التوصية

نظم التوصية(Recommender system) تمكن الناس من تبادل الآراء والإستفادة من خبرات بعضهم البعض ويمكن تعريفها بأنها أي نظام ينتج توصيات فردية كناتج له أوله تأثير على توجيه المستخدم بطريقة مخصصة لأشياء تثير إهتمامه أو أشياء مفيدة في مساحة واسعة من الخيارات المتاحة.

تم تطوير نظم التوصية في البداية لدعم مستخدمي الإنترنت في عملية صنع القرار في مواقف الحياة اليومية فيما يتعلق بإختيارهم للمعلومة التي قد تكون ذات فائدة لهم عندما يواجهون أوضاع ليس لهم فيها خبرة كافية في البدائل المتاحة .

٤.٢.٢أنواع نظم التوصية

و من أنواع نظم التوصية نذكر منها مايلي:

١.٤.٢.٢ الفلترة التعاونية (collaborative filtering)

هي توصيات تقوم على نموذج لسلوك مستخدم مسبق.نموذج يمكن بناؤه فقط من سلوك مستخدم واحد أو - على نحو أكثر فعالية - أيضا من سلوك المستخدمين الآخرين الذين لديهم صفات مماثله. عندما يأخذ سلوك المستخدمين الآخرين في الإعتبار، الفلترة التعاونية تستخدم معرفة المجموعة(group knowledge) لتشكيل التوصيات على أساس تشابه المستخدمين. وتعتمد التوصيات علي التعاون التلقائي لعدد من المستخدمين والمرشحين (automatic collaboration of multiple users and filtered) على أولئك الذين يظهرون تفضيلات أو سلوكيات مماثلة.

حقق نمو شبكة الإنترنت صعوبة أكثر لإستخراج المعلومات المفيدة من جميع المعلومات المتاحة على الإنترنت بشكل فعال.الكم الهائل من البيانات يتطلب آليات فعالة لتصفية المعلومات.ويطلق علي كل واحدة من التقنيات المستخدمة للتعامل مع هذه المشكلة الفلترة التعاونية (collaborative filtering). ودافع الفلترة التعاونية يأتي من فكرة أن الناس في كثير من الأحيان يحصلون على أفضل التوصيات من أشخاص مع أذواق مماثلة لأذواقهم، والفلترة التعاونية يستكشف التقنيات لمطابقة الناس

مع مصالح مماثلة وتقديم توصيات على هذا الأساس.وغالبا ما تحقق خوارزميات الفلترة التعاونيةالمشاركة الفعالة وطريقة سهلة لتمثيل مصالح المستخدمين على النظام.

الخوارزميات التي تكون قادرة على تطابق المستخدمين مع مصالح مماثلة،وواحدة من طرق نظم التوصية هي خوارزمية الجار الأقرب.

خوارزمية الجار الأقرب(K-Nearest Neighbor)

طريقة الجار الأقرب (K-NN) تعتبر من تقنيات التنقيب في البيانات التي تهدف للتنبؤ عن طريق مقارنة العناصر الشبيهة بالعنصر المراد التنبؤ به وتقدير القيمة المجهولة لهذا العنصر بناءاً على معلومات لتلك العناصر. وهي تقنية تنبؤية مناسبة لنماذج التصنيف،تصنف كائن من خلال تصويت الأغلبية من جيرانها،مع الكائن الذي يتم تعيينه إلى الطبقة الأكثر شيوعا بينها كأقرب جيرانها،إذ تمثل k عدد الحالات المتشابهة،وتعد معطيات التدريب في طريقة الجار الأقرب هي النموذج فلا يتم بناؤه عندما يتم تقديم حاله جديدة للبرنامج، تبحث الخوارزمية في المعطيات كلها لإيجاد علاقة جزئية من الحالات التي هي أكثر تشابهاً بها، وتستخدمها لتوقع الخرج.

محددات خوارزمية الجار الأقرب

١.عدد الحالات الأقرب ليتم استخدامها(k).

٢.وحدة قياس التشابه.

يتطلب كل إستخدام لخوارزمية الجار الأقرب أن يتم تحديد قيمة موجبة ل k،وهذا يحدد عدد الحالات الموجودة التي يتم البحث عنها عند تحديد حالة جديدة للخوارزمية.

١.١.٤.٢.٢ أنواع الفلترة التعاونية

الفلترة التعاونية[6] لديها ثلاثة أنواع هي:-

- **القائم علي الذاكرة (Based on memory)**

يستخدم هذا النهج بيانات تقييم المستخدم لحساب التشابه بين المستخدمين أو العناصر، ويستخدم لتقديم التوصيات.كان هذا النهج في وقت مبكر يستخدم في كثير من النظم التجارية.وهو فعال وسهل التنفيذ.أمثلة نموذجية لهذا النهج هي(neighborhood-based CF) و (item-based/user-based) top-N.

ومن مزايا هذا النهج توضيح النتائج وهو جانب هام من جوانب نظم التوصية؛ إنشاء سهولة الإستخدام؛ تيسير البيانات الجديدة بسهوله؛ إستقلال المحتوى من العناصر التي يوصى بها وتوسيع نطاق جيد مع تصنيف العناصر المشتركة.

هناك أيضا العديد من العيوب مع هذا النهج.ينخفض أدائها عندما يحصل على البيانات المتفرقة، والذي يحدث في كثير من الأحيان مع العناصر ذات الصلة على شبكة الإنترنت، وهذا يعيق قابلية هذا النهج ويخلق مشاكل مع مجموعات البيانات الكبيرة. على الرغم من أنه يمكن التعامل مع المستخدمين الجدد بكفاءة لأنه يعتمد على بنية البيانات فإن بإضافة عناصر جديدة يصبح أكثر تعقيداً، وإضافة عناصر جديدة يتطلب إدراج جديد وإعادة الإدراج لجميع العناصر في البنية.

- **القائم علي النموذج (model Based on)**

يتم فيه وضع نماذج بإستخدام التنقيب عن البيانات، بإستخدام خوارزميات تعلم الآلة لإيجاد أنماط تستند إلى بيانات التدريب. وتستخدم هذه لجعل التوقعات للبيانات حقيقية.وتستند معظم النماذج على خلق أسلوب تصنيف أو تجميع لتعريف المستخدم إستناداً إلى مجموعة الإختبار.هناك العديد من المزايا مع هذا النموذج يتعامل بها مع تبعثر أفضل من القائم على الذاكرة. وهذا يساعد على تطويره مع مجموعات كبيرة من البيانات، وهو يحسن أداء التنبؤ، إذ أنه يعطي المبرر بديهياً للتوصيات.

- **المختلط (Hybrid)**

يجمع بين القائم على الذاكرة والقائم على النموذج، وهو يحسن أداء التنبؤ، الأهم من ذلك أنه يتغلب على مشاكل (CF)مثل تبعثر وفقدان المعلومات. ومع ذلك، فإنه زاد من التعقيد وتكلفة التنفيذ.عادة معظم نظم التوصيةالتجارية هي مختلطةمثل، (Google news recommender system).

٢.١.٤.٢.٢ تحديات الفلترة التعاونية

يوجد العديد من التحديات التي واجهة الفلترة التعاونية [6]منها:

- **تبعثر البيانات (Scattering Data)**

في الممارسة العملية، تقوم العديد من نظم التوصيةالتجارية على مجموعات البيانات الكبيرة، وهي كبيرة للغاية ومتفرقة، وهذا الذي يجلب التحديات في أداء هذه التوصية. ويوجد مشكلة واحدة نموذجية ناجمة عن تبعثر البيانات هي مشكلة(cold start) (الأنظمة التي توجد بها cold start غالبا ما تتطلب كمية كبيرة من البيانات للمستخدم من أجل تقديم توصيات دقيقة).

- **قابلية التوسع (Scalability)**

أن أعداد المستخدمين و البنود تنمو بسرعه كبيره، وخوارزميات (CF) التقليدية تعاني من مشاكل (Scalability) الخطيرة وكذلك تحتاج العديد من الأنظمة للرد على الفور لمتطلبات الإنترنت وتقديم توصيات لجميع المستخدمين بغض النظر عن مشترياتهم ودرجات التقييم، والتي تتطلب قابلية أعلى من نظام (CF).

- **المرادفات(Synonyms)**

المرادفات تشير إلى اتجاه عدد من البنود ذاتها أو مماثله لها التي لها أسماء أو إدخالات مختلفة. على سبيل المثال، على ما يبدو عناصر مختلفة "فيلم الأطفال" و "سينما الأطفال" يحيل في الواقع إلى نفس البند،ودرجة التباين في استخدام مصطلح وصفي أكبر مما كان يعتقد عادة. وانتشار المرادفات يقلل من أداء أنظمة التوصية (CF). وبأسلوب النمذجة (مثل تقنيةDirichlet Allocation) يمكن حل ذلك عن طريق تجميع كلمات مختلفة تنتمي إلى نفس الموضوع.

- **اعتداءات (SHILLING)**

في نظام التوصية حيث يمكن لأي شخص أن يعطي تقييمات،والناس قد تعطي الكثير من التصنيفات الإيجابية للعناصر الخاصة بهم ودرجات تقييم سلبية على منافسيهم. فمن الضروري في كثير من الأحيان لأنظمة (CF) تقديم الإحتياطات للحد من هذا النوع من التلاعب.

١.٤.٢.٢فلترةعلى المحتوى (Content-based filtering)

الفلترة على المحتوى[7](Content-based filtering)يشار لها أيضاً بالتصفية المعرفية (cognitive filtering)،وتوصي بالعناصر بناءاً علي مقارنة بين محتوي العناصر وملف تعريف المستخدم. ويمثل محتوي كل عنصر أنه مجموعة من أوصاف أو مصطلحات. وملف تعريف المستخدم يمثل بنفس الشروط ويتم بناؤه من خلال تحليل محتوى العناصر التي قد يراها المستخدم.

مصدر المعلومات الذي يستخدم أنظمة الفلترة على المحتوى في الغالب مع (Text documents)وثائق النصهونهج موحد لتحليل المدى الذي يتم فيه إختيارعبارة واحدة من الوثائق.

نموذج ناقلات الفضاء(The vector space) وفهرسة الدلالات الكامنة latent semantic indexing)الطريقتان اللتان تستخدمان لتمثيل الوثائق.

١.٤.٢.٢ نظم التوصية المختلطة(Hybrid Recommender System)

وقد أثبتت الأبحاث الحديثة أن إتباع النهج المختلط[7]يجمع بين الفلترة التعاونية والفلترة علي المحتوى ويمكن أن يكون أكثر فعالية في بعض الحالات. ويمكن تنفيذ النهج المختلط بعدة طرق: عن طريق جعل تنبؤات الفلترة علي المحتوى والفلترة التعاونية كلٍ على حدة ومن ثم الجمع بينهما؛ بإضافة قدرات على الفلترة علي المحتوي للنهج القائم على الفلترة التعاونية(والعكس بالعكس)،أوعن طريق توحيد النهج في نموذج واحد.

مثال جيد على الأنظمة المختلطه هو (Netflix)، يعطي التوصيات بمقارنة و مراقبة وتفتيش عادات المستخدمين المماثلة (أيCF)، وكذلك من خلال تقديم العناصر التي تشترك مع خصائص العناصر التي قام مستخدم بها بدرجات عالية (content-based filtering).

يستخدم نظام (hybrid recommender) لوصف أي نظام توصية يجمع بين تقنيات توصية متعددة معاً لإنتاج المخرجات الخاص به.لا يوجد أي سبب يمكن أن يمنع (hybrid)أن يعمل بعدة تقنيات مختلفة من نفس النوع، على سبيل المثال، يمكن لإثنين من أنواع الفلترة علي المحتوى المختلفة أن تعمل معاً،وحققت عدداً من مشاريع هذا النوع (hybrid): (News Dude)، والذي يستخدم كلاً naive (Bays) ومصنفات (KNN)في توصيات أخبارها.

الفصل الثالث

أدوات تنقيب البيانات

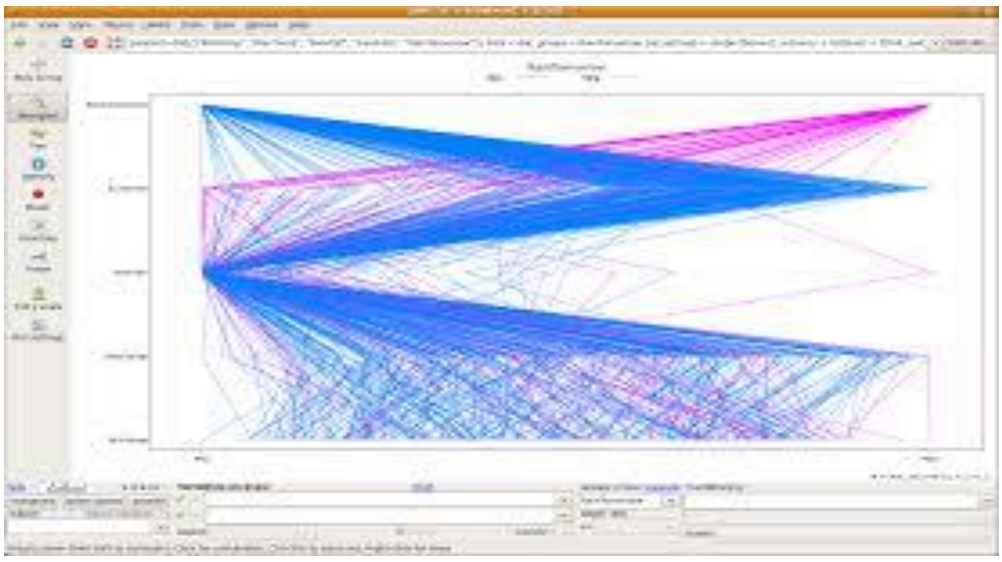
١.٣.٢مقدمة

تطوير وتطبيق خوارزميات التنقيب عن البيانات يتطلب إستخدام أدوات برمجية قوية. كما لا يزال عدد من الأدوات المتاحة تتزايد في النمو، مما يصبح إختيار الأداة الأكثر ملائمة من أكثر الأمور صعوبة، وفيما يلي سنتطرق لعدد من أدوات تنقيب البيانات.

Rattle٢.٣.٢

هي البرمجيات الحرة والمفتوحة المصدر وحزمة توفير واجهة المستخدم الرسومية (GUI) لإستخراج البيانات بإستخدام لغة البرمجة الإحصائية R. يستخدم (Rattle)[8] في مجموعة متنوعة من الحالات. يقدم (Rattle)ملخصات إحصائية ومرئية من البيانات مع (2D) و (3D)، ويوفر وظائف كبيرة لإستخراج البيانات من خلال تعريض قوة البرنامج الإحصائي R خلال واجهة المستخدم الرسومية. كما يستخدم كمنشأة تعليمية لتعلم لغة البرمجة R.

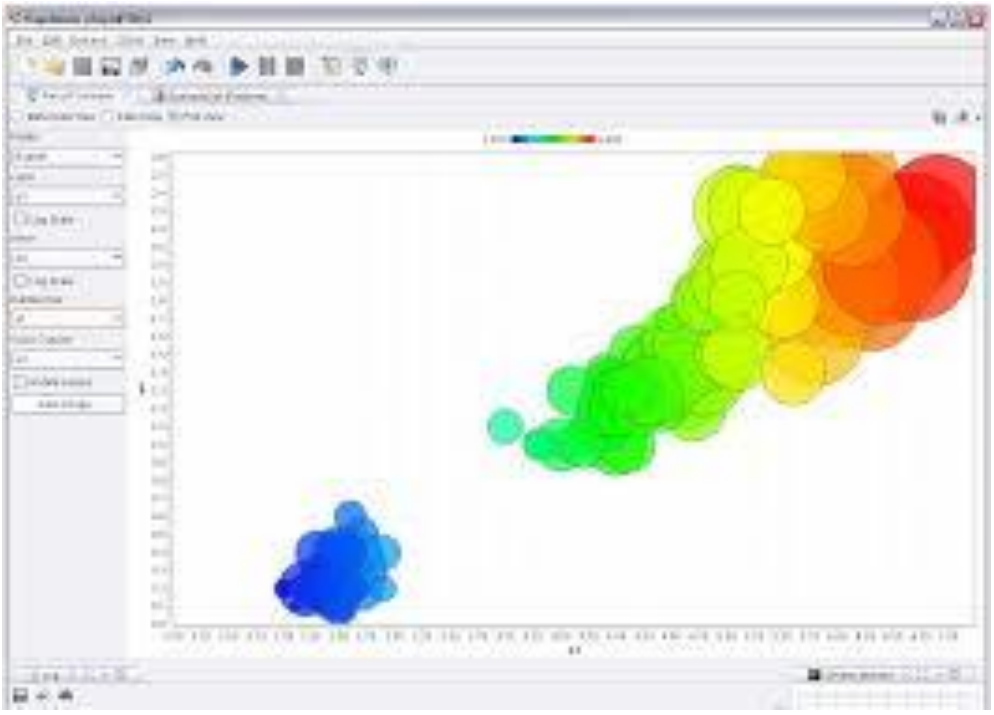
يحول البيانات التي يمكن أن تكون منمذجة(model) بسهولة، في النموذجين(unsupervised) و(supervised)، أما الغير خاضع للرقابة (unsupervised)يستخدم طريقة التحليل العنقودي أو قواعد الإرتباط، والخاضع للرقابة (supervised)يستخدم لتوقع الأنماط المستقبلية للسلوك من البيانات، ويقدم أداء النماذج بالرسوم والصور، ونتائج مجموعات البيانات الجديدة. ويمكن إستخدامه لأغراض التحليل الإحصائي، أو توليد نموذج. و يسمح لمجموعة البيانات أن تقسم في التدريب، والتحقق والإختبار.



شكل(٢.٢)واجهة ال Rattle

Rapid Miner ٣.٣.٢

بالإضافة إلى استخراج البيانات يوفر[9](Rapid Miner)وظائف مثل تجهيز البيانات والتصور، والتحليلات التنبؤية والنمذجة الإحصائية والتقييم والنشر. مما يجعلها أكثر قوة هو أنه يوفر التعلم بالمخططات والنماذج والخوارزميات من البرامج النصية .



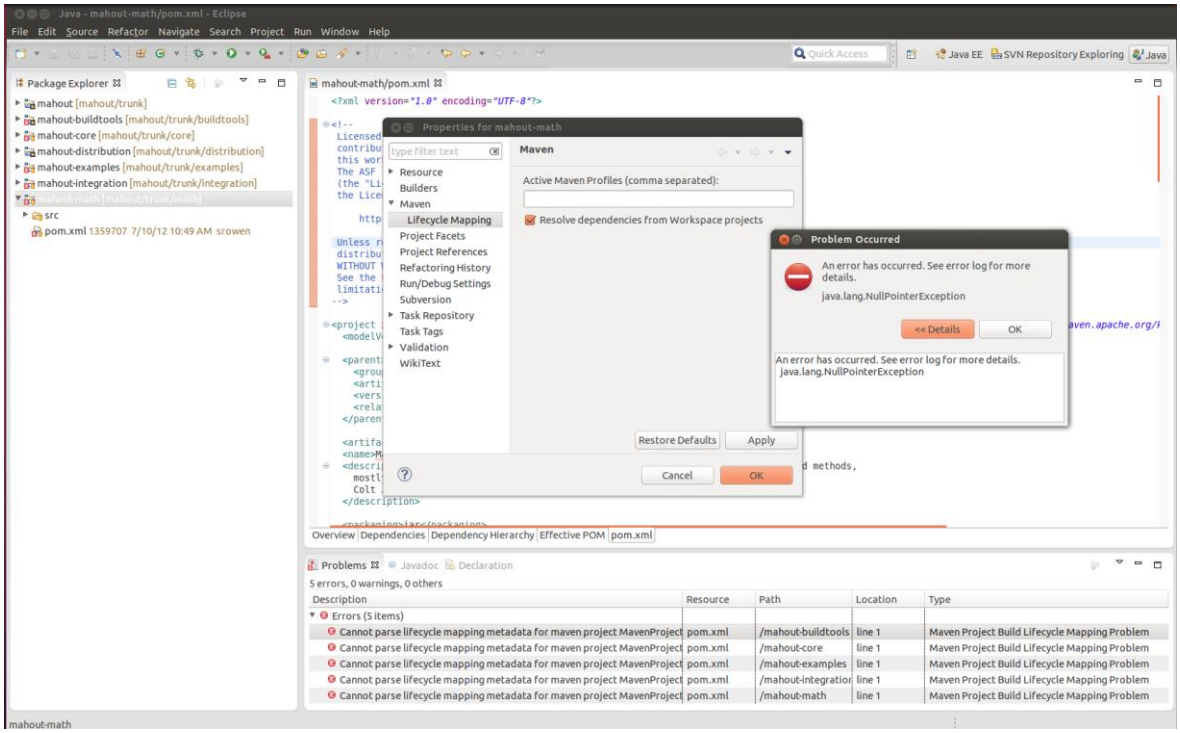
شكل(٣.٢)واجهةRapid Miner

Apache Mahout ٤.٣.٢

يقدم (Apache Mahout) [10] بيئة الرياضيات الجديدة التي تسمى (Samsara). شعاره التجديد الشامل، وهو يعكس إعادة التفكير بشكل أساسي .والقابلية لبناء خوارزميات تعلم الآلة حسب الطلب. وهو يساعد الناس على خلق رياضيات خاصة مع توفير بعض حالات التنفيذ الجاهزة للخوارزمية.

آلة تعلم مفتوحة المصدر، يلتقط الخوارزميات الأساسية لنظم التوصية، والتصنيف، والمجموعات في المكتبات الجاهزة القابلة للاستخدام، وهو يدعم تنقيب التوصيات (recommendation mining)، ومع (Mahout) يمكن أن يطبق تطبيق مباشر إلى المشاريع الخاصة بتقنيات التعلم الآلي التي توجه الأمازون، ونيتفليكس، وغيرها.في جوهرها هي الجبر الخطي العام والعمليات الإحصائية إلى جانب هياكل البيانات لتقديم الدعم لهم، وجعل النماذج أو مهمة التقديم أسهل بكثير وسمح للمستخدمين تخصيص

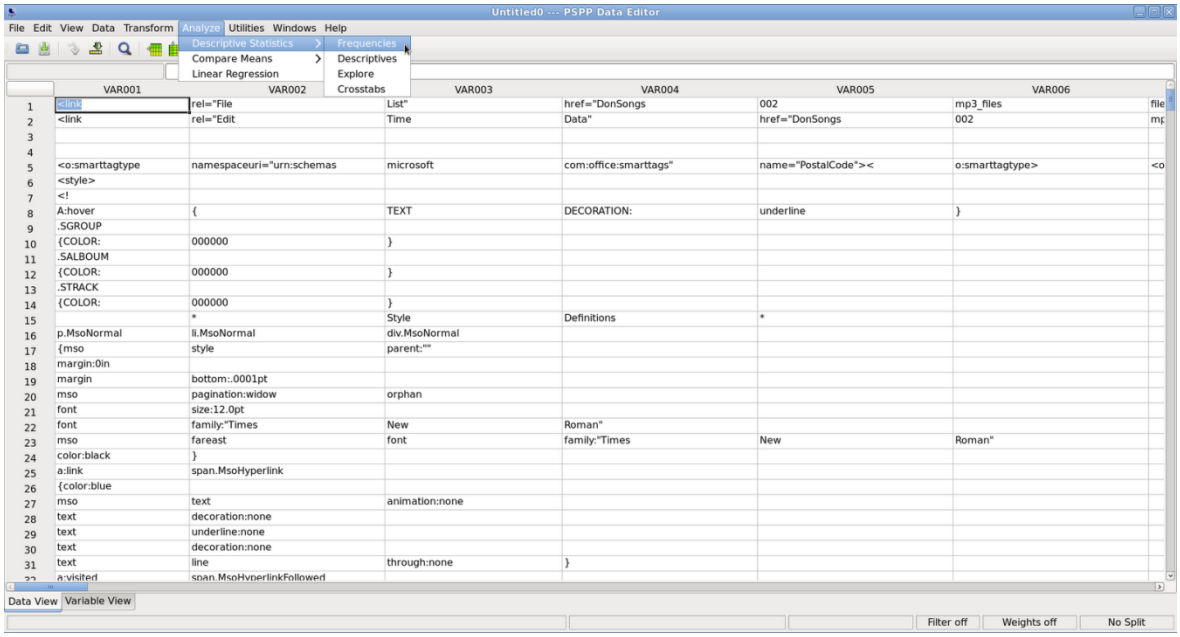
الخوارزميات مع درجة جديدة كاملة من الحرية.تشمل خوارزميات(Mahout) العديد من التطبيقات الجديدة التي بنيت للسرعة، والسرعة فية زادت بمقدار ١٠ أضعاف عن الخوارزميات التي سبقتها



شكل(٤.٢)واجهةApatch Mahout

PSPP ٥.٣.٢

PSPP هو برنامج للتحليل الإحصائي لبيانات العينة. أنه يحتوي على واجهة المستخدم الرسومية والتقليدية. مكتوب في C، ويستخدم (plotutils) لتوليد الرسوم البيانية. وهو بديل مجاني لبرنامج الملكية SPSS (منIBM)، ويتنبئ بثقة لما سيحدث ثم بعد ذلك يمكنك من إتخاذ قرارات أكثر ذكاءاً، ويقدم حلاً للمشاكل وتحسين النتائج[11].

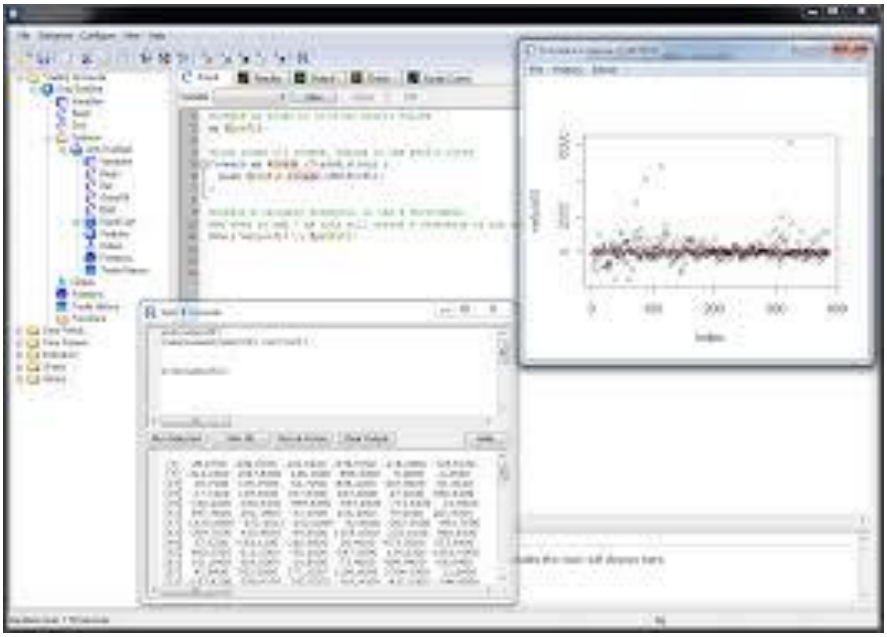


شكل(٥.٢)واجهة Pspp

R ٦.٣.٢

هي لغة وبيئة الحوسبة الإحصائية والرسوم البيانية، يتم إستخدام لغةR[12]على نطاق واسع بين تعدين البيانات لتطوير البرامج الإحصائية وتحليل البيانات، وأثار سهولة الإستخدام والتمدد. يوفر R مجموعة واسعة من الإحصائية (النمذجة الخطية وغير الخطية ، والإختبارات الإحصائية الكلاسيكية، وتحليل السلاسل الزمنية، والتصنيف، والتجميع، ...) والتقنيات الرسومية.

واحدة من نقاط قوة Rهي السهولة التي يمكن أن تنتج تصميماً جيداً ، بما في ذلك الرموز الرياضية والصيغ. وقد تم الحرص الشديد على الإعدادات الإفتراضية لخيارات التصميم الثانوية في الرسومات، ولكن يحتفظ المستخدم بالسيطرة الكاملة.



شكل(٦.٢)واجهةR

Weka ٦.٢

هو عبارة عن برنامج خاص بخوارزميات التنقيب عن البيانات وتعلم الآلة[13]. كما يحتوي هذا البرنامج على أدوات قادرة على التعامل مع الأمور التالية :

(pre-processing)،(classification)،(clustering)،(association rules)،(visualization)

و يمكن إستخدامه من خلال الواجهة الرسومية للتفاعل مع ملفات البيانات وتحقيق نتائج بصرية، وهو يعمل تحت عدة بيئات (جنو\لينوكس، ويندوز، ماك). ليعمل الملف يجب أن يكون في صيغ معينه ومن هذة الصيغ صيغة (arff).

١.٧.٣.٢ملف (ARFF)

لتحويل الملف إلي الصيغة (arff) يجب حفظ ملف (excel) إلي الصيغة (csv) أولاً. (Arff) هي صيغ تكون في شكل صفوف يتم فيها توضيح السمات الأساسية للملف أعلي التعليقات وهي يجب أن تكون مسبوقة بالعلامة@ ويمكن أن تحتوي السمات علي فراغات،السمات الإسمية تكون (string)والرقمية (numeric).

مثال لصيغ ملف (arff) للمواد (material)

@attribute security numeric

@attribute os numeric

@attribute ai numeric

@attribute hci numeric

@attribute physics1 numeric

@attribute hardware numeric

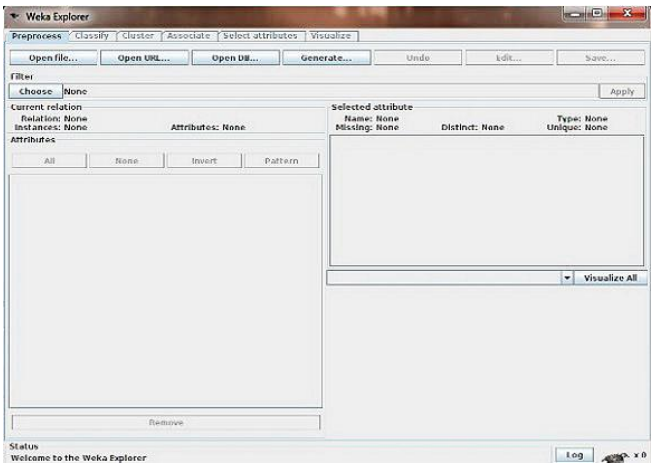
@attribute math1 numeric

@attribute Entrance PC

٢.٧.٣.٢ تحميل البيانات في weka

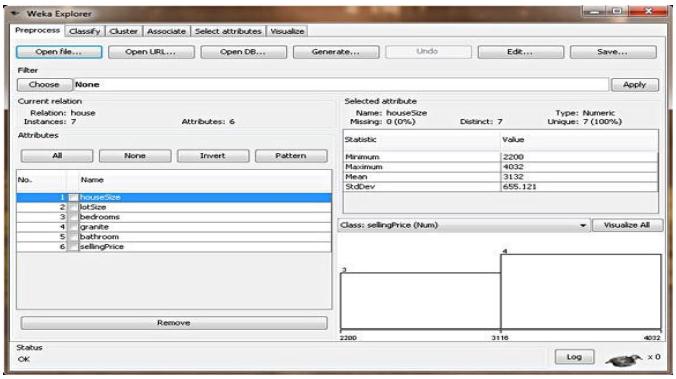
الشكل (٧.٢) شاشة بدء تشغيلWEKA

عند بدء تشغيل Weka تظهر في البدأ واجهة الإختيار الرسومية تتيح لك إختيار أربع طرق للعمل معWekaوالبيانات الخاصة بك، وسوف نختار خيار واحد والخيار(explorer) أكثرمن كافي لكل ما نحتاج للقيام به في هذه البيانات.وعند إختيار إكسبلورر تفتح شاشة المستكشف الرسوميةوالشكل يوضح شاشة المستخدم قبل بدء الإختيار.



الشكل (٨.٢) شاشة المستكشف الرسومية قبل الإختيار

لتحميل البيانات إلى WEKA، يجب وضعها في الشكل الذي سوف يكون مفهوماً. الأسلوب المفضل لتحميل البيانات أن يكون الملف (ARFF)، في البدء إختيار (open) لإختيار الملف ذو الإمتداد (ARFF)الذي من خلاله يمكن أن تختار الملف الذي تريد تحميله. ثم إختيار (apply) بعد تحديد نوع البيانات التي يتم تحميلها من الخيار (choose) لتعمل الخوارزمية. الشكل يوضح weka بعد إختيار البيانات.



الشكل (٩.٢) شاشة المستكشف الرسومية بعد الإختيار

الباب الثالث

الدراسات السابقة

٣.١مقدمة

أثبتت نظم التوصية النجاح في العديد من المجالات التي توجد فيها مشكلة الحمل الزائد من المعلومات(information overload).وعلى الرغم من أن أنظمة التوصية التعليمية (Recommendation educational systems) تشترك في نفس الأهداف الرئيسية مع التوصية لتطبيقات التجارة الإلكترونية، هنالك بعض الخصائص التي ينبغي النظر فيها قبل تطبيق الحلول الحالية مباشرة من تلك التطبيقات. على سبيل المثال، لا ينبغي أن تسترشد التوصيات في المجال التعليمي فقط من خلال تفضيلات المتعلمين ولكن أيضا من خلال المعايير التعليمية[14].

٣.٢دراسة عن تطبيق التوصية المختلطة في التعليم الالكتروني

E-Learning personalization Based On Hybrid Recommendation Strategy And Learning Style Identification:

من المهم جداً توفير نظام مخصص يمكنه أن يتكيف تلقائياً مع أساليب التعلم للمتعلمين ومستويات معرفتهم وذكاءهم، يوصي بالأنشطة عبر الإنترنت أو المواد التي يفضلونها ويحسن التعلم.في سياق التعليم اعتبر الباحثين[15].أن المتعلمين يمكن أن يكون لديهم إحتياجات فردية مختلفة وخصائص مختلفة مثل مستويات مختلفة من الخبرة والمعرفة السابقة، والقدرات المعرفية، وأساليب التعلم، والتحفيز، وأنهم يريدون تحقيق إختصاص معين في وقت معين. وبالتالي، فإنه لا يمكن أن يعاملوا بطريقة موحدة، يسمي علماء النفس هذه الفروق المختلفة بأنماط التعلم.

تم تقديم نظام توصية ذكي للتكيف على شبكة الإنترنت وهو بروتيوس (نظام برمجة الدروس الخصوصية) لمساعدة المتعلمين في تعلم أساسيات لغات البرمجة. والذي يمكنه التكيف تلقائياً مع مصالح ومستويات المعرفة للمتعلمين، ويأخذ بعين الإعتبار الجوانب التربوية للمتعلم وضرورة أن يوصي بسلاسل من أنشطة التعلم بترتيب فعال تربوياً.

من أجل التحقيق في أنشطة التعلم بالتفصيل، إعتمدت خوارزمية تنقيب أنماط متتابعة Apriori) All) (Tong & Pi-Lian, 2005) لإستخراج الأنماط (التفاعل) السلوكية من ملف السجل. وهذه الأنماط تكون مفيدة لتحليل كيفية تطور المتعلمين من بداية تعلم وحدة معينة. المتعلمين الذين لديهم أساليب تعلم مختلفة لديهم مجموعات مختلفة من سلاسل مختلفة. وبالتالي، تم تجميع المتعلمين في مجموعات على أساس أساليب التعلم الخاصة بهم ثم تم إكتشاف الأنماط السلوكية لكل متعلم بواسطة

خوارزمية (Apriori All). أما مهمة نظام التصفية التعاونية هي التنبؤ بتقييم إستفادة متعلم معين مقارنة بمتعلم مماثل.

إستخدموا واحدة من أدوات جمع البيانات تسمى مؤشر أساليب التعلم (ILS) (Felder وSoloman ، وILS هو ٤٤ سؤال، وهو متاح مجانا، وسيلة إختيار عدة أساليب و الذي يقيم الإختلافات الفردية في تفضيلات أسلوب التعلم.

قبل مرحلة معالجة البيانات لبناء نموذج المتعلم، يتم جمعالبيانات حول النشاطات (مثل الأنماط التتابعية، الصفحات التي زارها، نتائج الاختبارات، والدرجات التي تحصل عليها). صفحات تسجيل المتعلمين. الجزء النظري للدورة والدروس.الامثلة والاختبارات من معالجة البيانات المدخلة.

نفذوا بعض التجارب على مجموعة البيانات التعليمية. إختاروا ٤٤٠ من المتعلمين، في مرحلة التخرج من قسم تكنولوجيا المعلومات في المدرسة العليا للدراسات الأعمال الفنية، Novi Sad University.المتعلمين المشاركين يضموا مبرمجين مبتدئين إجتازوا بنجاح الدورة الأساسية لمحو الأمية المعلوماتية في الفصل الدراسي السابق.وتم تقسيمهم إلى مجموعتين: المجموعة التجريبية والمجموعة المسيطرة.

المتعلمين من المجموعة المسيطرة عملت بطريقة طبيعية ولم تتلق أي توصية أو توجيه من خلال مسارها التعليمي، في حين أنه كان المطلوب من طلاب المجموعة التجريبية إستخدام نظام بروتيوس.

المتعلمين من كلا الفريقين لم يتخذوا أي مسار تقليدي موازي وطلب منهم عدم إستخدام أي مواد إضافية أو مساعدة.

وتألفت المجموعة التجريبية من ٣٤٠ متعلم، بينما تألفت المجموعة الضابطة من ١٠٠ متعلم. إستمرت التجربة لمدة أربعة أشهر تقريبا، خلال الفترة من يناير ٢٠١٠ وحتى أبريل ٢٠١٠.إستنتجوا أن الإختلافات بين المجموعتين التجريبية والمجموعة المسيطرة تكاد لا تذكر.

أظهرت نتائج التجربة أن المتعلمين في المجموعة التجريبية يجب أن يكونوا قادرين على إتمام الدورة في وقت أقل من المتعلمين في المجموعة المسيطرة. وأن المجموعة التجريبية أنهت بإستمرار المزيد من الدروس بنجاح أكثر من المجموعة المسيطرة.

كما أظهرت النتائج ملائمة إستخدام هذا النموذج للتوصية، من أجل اقتراح أنشطة التعلم عبر الإنترنت للمتعلمين على أساس أسلوب التعلم والمعرفة والتفضيلات. ولقياس الدقة الإحصائية إستخدم متوسط الخطأ المطلق (MAE).

٣.٣مزيج من خوارزميات تعلم الآلة للتوصية بالمقررات في نظام التعليم الإلكتروني اعتمادا على البيانات التاريخية:

Combination Of machine learning Algorithms For Recommendation Of Courses In E-Learning System Based On Historical Data:

في نظام التوصية للمقررات التي توصي المقررات للطالب على أساس إختيار الطلاب الآخرين لمجموعة معينة من المقررات[16]. ونتيجة لنظام التوصية للمقررات، يمكن أن يوصي لطالب جديد لم يلتحق مسبقاً بأي مقرر. إستخدموا في طريقتهم مزيج من أسلوب التجميعات (Simple K-means) وخوارزمية(Apriori)، لإيجاد النتيجة. وتمت مقارنة هذه النتائج مع نتائج أداة تنقيب البيانات Weka المفتوحة المصدر.

قاموا بمقارنة نتيجة خوارزمية قواعد الإرتباط (Apriori) وهي خوارزمية موجودة في أداة تنقيب البيانات weka مفتوحة المصدر ومجموعة أخرى من خوارزميات التجميع وقواعد الإرتباط مع نتيجة خوارزمية الجمع بإستخدام تجميع (K-means) وخوارزمية قواعد الإرتباط (Apriori).

أخذ في الاعتبار فئتين، فئة ١٣ مقرر وفئة ١٨ مقرر التي تخص قسمي علوم وهندسة الحاسوب (CSE) وتكنولوجيا المعلومات (IT). بعد جمع البيانات يهتمون فقط لأولئك الطلاب الذين يلتحقون بما لا يقل عن ٥ دورات.

النتيجة بإستخدام خوارزمية (Apriori) تحتاج مرحلة إعداد البيانات وكذلك إذا زدنا الدعم فان عدد قواعد الإرتباط سيكون أقل. و النتائج بإستخدام مزيج من مجموعات التجمعيات الأخري لا يتطابق مع الخيارات في العالم الحقيقي للطلاب ولكن إذا إستخدموا طريقة مشتركة بإستخدام (K-means) و(Apriori) فعندها ليس هناك حاجة لإستخدام مرحلة إعداد البيانات وعدد قواعد الإرتباط سيكون أكثر، هذه الطرق المشتركة ستزيد قوة قواعد الإرتباط لذلك نظام التوصية للمقررات هذا سيساعد الطالب ليختار المقرر المناسب وفقاً لرغبتة.

تم عرض النتائج بإستخدام Wekaإلي جانب نتائج الطريقة المشتركة (Simple K-means) و(Apriori)، ووجد أنها أفضل من النتيجة التي يتحصلوا عليها بإستخدام الخوارزميات والتجميعات الأخري من مجموعات أخرى من خوارزميات التجميع وقواعد الإرتباط .

هذا النظام يمكن إستخدامه أيضا ليوصي بالمقررات للطالب في التعلم عن بعد يكون العمل في المستقبل لإيجاد مزيج أخر من تقنيات إستخراج البيانات التي قد تكون مزيجاً من التصنيف وخوارزميات قواعد

الإرتباط أو مزيج من التصنيف والتجميع وخوارزميات قواعد الإرتباط. هذا العمل يمكن أيضاً أن يكون أكثر توسعاً وتطوراً بحيث يدمج مع نظام التعليم الإلكتروني الموجود. لذلك نظام التوصية للمقررات هذا سيساعد في بناء نظام توصية ذكي. هذه الطريقة للتوصية للمقررات للطلاب الجدد تستخدم بشكل كبير في (MOOC (Massively Open Online Courses).

٣.٤النهج القائم على تصفية التعاونية لتوصية المقررات الاختيارية:

A Collaborative Filtering Based Approach for Recommending Elective Courses

التنبؤ بالدرجة يكون مفيد جداً للطالب ليقوم بإختيار المواد التي يرغب بها وقد تناسبه، وفي الوقت نفسه يتحصل فيها على درجات أفضل نسبياً. ويواجه الطلبة الذين يسعون إلى درجة التعليم العالي تحديين: عدد لا يحصى من الدورات للإختيار من بينها، وعدم وجود الوعي بالدورات التي يجب أن يدروسونها وبأي ترتيب. فإنه وفقا لأصدقائهم وتوصيات زملاءهم فأن الكثيرين منهم يختاروا مقرراتهم ويسجلوا وفقا لذلك.

تم تقديم نظام التوصية [17].لإختيار المقررات الدراسية الإختيارية ودمجه في نظام التسجيل الحالي لكلية إدارة الأعمال إما عبر الإنترنت أو بدونه. ويستند نظام التزكية للمقررات على عدة خوارزميات الفلتره التعاونية المختلفة مثل (a modified variant of item - based OC1), (user- based C4.5). وهي تقدم توصيات العنصر أو التوقعات إستناداً إلى آراء المستخدمين الآخرين المشابهين للمستخدمين الحاليين في التفكير. كما أنها تتطلب الحد الأدنى من المعلومات، وهي عالية الدقة في التوصيات.

تمت مقارنة نهجين للتصفية التعاونية للتنبؤ بدرجة الطلبة التي قد يتحصلوا عليها في الدورات المختلفة، بناءاً على أدائهم في الدورات السابقة. والفرق الأساسي بين الطريقتين،أن (item_based) تنتج التنبؤات إعتماداً علي نموذج تشابه عنصر-لعنصر أما (user_based) تعتمد علي أساس نموذج تشابه مستخدم –لمستخدم وتستخدم أساساً لحل مشكلة التنبؤ. ولحساب التشابه تم إستخدام معامل بيرسون.

تصنيفات البيانات يتحصل عليها صراحة أو ضمنياً. صراحة هي تلك التي تعطى تقييمات صريحة مباشرة من قبل المستخدم، في حين أن البيانات الضمنية يمكن جمعها بشكل غير مباشر من خلال دراسة بيانات عن المستخدم من مصادر مختلفة مثل بيانات الشراء، وإستعراض السلوك الخ.

نتائج التجريب لديهم على مجموعة البيانات من الحياة الحقيقية مشجعة للغاية. وهم يعتقدوا أن النهج الذي إتبعته التصفية التعاونية الطلابية للتنبؤ بالدرجة، دقيق. ومن النتائج وقيم MAE تمت ملاحظة أنه لا يوجد فرق كبير في الدقة بين (user_based) و(item_based).

للحصول على توصيات دقيقة، لا بد من تقييم العديد من المقررات قدر الإمكان. النظام لا يمكنه التنبؤ بتوصيات للطلاب الذين لم يدرسوا أي من المقررات الدراسية في الجامعة.للعمل في المستقبل، يمكن أن يتم البحث في وضع استراتيجيات التكامل للأساليب التي يمكنها التنبؤ بدقة لأداء الطلاب في الدورات والمناهج التي تساعد على اختيار مواضيع أو دورات على أساس إهتمامات الطلاب وأهداف التعليم. هذه الطرق يمكن إستخدامها لتقديم النصائح القيمة للطلاب وتقديم المشورة والتوجيه المهني وعملية إختيار المقررات.

٥.٣ استخدام تنقيب البيانات لتحسين استبقاء الطالب في التعليم العالي

Use Data Mining To Improve Student Retention In Higher Educaction

واحدة من أكبر التحديات التي يواجهها التعليم العالي هو تحسين معدلات إستبقاء الطالب لمواصلة تعليمه. يستخدم الباحثين[18]. إستخراج البيانات وتقنيات معالجة اللغة الطبيعية لمراقبة الطلاب، وتحليل السلوك الأكاديمي للطلاب وتوفير أساس لإستراتيجيات تدخل تتسم بالكفاءة. أن وظيفة التنبؤ من إستخراج البيانات يمكن أن تساعد الجامعة لإستبقاء الطلاب أو التخطيط لموارد تستند إلى معرفة كم عدد الذين سينتقلوا أو سيسجلوا لدورة معينة. هدفهم هو تحديد المشاكل المحتملة في أقرب وقت ممكن ومتابعتها مع خيارات التدخل لتحسين استبقاء الطالب.

اكتشاف المعرفة في قواعد البيانات (KDD)، ويمكن العثور علي العلاقات والأنماط الموجودة لكنها مخبأة بين الكم الهائل من البيانات التعليمية. فهو يجمع بين التعلم الآلي والتقنيات الإحصائية والتصور لاكتشاف واستخراج المعرفة بمثل هذه الطريقة يمكن للناس فهمه بسهولة. للجامعات، فإن المعرفة التي اكتشفت بواسطة تقنيات استخراج البيانات توفر تعليم مخصص يرضي مطالب الطلاب واصحاب العمل.

نظام جامعة وادي التايمز (Oizilbash) (TVU، 2008) لديها كمية كبيرة من البيانات التي يمكن تحليلها واستخراجها لنظام استخراج البيانات. الكليات والأقسام لديها بيانات التفاصيل الهامة أيضا بشأن الدورات والوحدات وهي في شكل وثيقة. نظام سجل الطالب يحمل معلومات حول سجلات الطلاب

سبيل المثال، نتائج الامتحانات للطالب والتسجيل للدورات. وهو مصدر البيانات الأكثر أهمية لمشروعهم.

(Seidman,1996) وضع سيدمان صيغة لاستبقاء الطالب:

Retention = Early Identification + (Early + Intensive +Continuous) Interventio

يقومون الان بتطوير جزء متداخل من MCMS يحتوي علي موقع علي الويب لمراقبة الطالب،والوحدة، ومعلومات الدورة عندما يسجل بها الطالب، كما انه يمكن ان يتحقق من المعلومات التفصيلية لكل نتائج النماذج ومخطط تطوره وما اذا كان في خطر ام لا ولماذ.اما بالنسبة للمدرسين فيمكنهم الدخول لايجاد تفاصيل متعلقة بمعلومات الوحدة، ويمكن ان يتواصلوا مع الطالب عبر النظام، يمكنهم أيضا مقارنة المعلومات من وحدات ذات صلة في فصول دراسية مختلفة.

الباب الرابع

الإرشاد الأكاديمي

٤.١مقدمه

يعد الإرشاد الأكاديمي الركيزة الأولى التي يعتمد عليها نظام الساعات المعتمدة وهو لا يقتصر في وظائفه على التعريف بالخطة الدراسية ومساقاتها فقط، بل يتعداها إلى توثيق العلاقة بين الطلبة وأولياء الأمور من جانب وأعضاء هيئة التدريس من جانب آخر، [19]بما يحقق الأهداف المتوحاة من العملية التربوية.وهو عملية مهنية وتربوية تقدم للطلبـة الدارسين في الجامعات. ويمثل الإرشاد الأكاديمي واحداً من أهم وسائل مساعدة الطالب للإندماج في البيئة الأكاديمية وتحقيق أعلى معدلات النجاح، كما يساهم في التأكد من مسيرة الطالب العلمية من حيث تطبيق الخطة الدراسية وكذلك فيما يتعلق بأداء الطالب وتحصيله العلمي في مختلف المساقات. وأيضا يساهم الإرشاد الأكاديمي في مساعدة الطالب على الإنتقال التدريجي والمبرمج من بيئة التعليم في المراحل السابقة لدراسته الجامعية إلى بيئة المرحلة الجامعية التي يعتمد فيها الطالب على نفسه في إتخاذ قراراته وتحديد تخصصه وتطوير مستواه العلمي والسلوكي. وتقوم الدوائر المختلفة بتعيين مرشد أكاديمي لكل طالب منذ إلتحاقه بالجامعة بالفصل الاول.

٢.٤ الإرشاد الأكاديمي

هو النشاط الذي يقوم به المرشدون الأكاديميون بالكلية لتعريف الطلاب بالأنظمة واللوائح الدراسية ومساعدتهم في التقدم الدراسي والتغلب على ما يعترضهم من مشكلات وصعوبات إجتماعية أو نفسية أو صحية أو أكاديمية.

٣.٤ أهمية الإرشاد الأكاديمي

الطالب هو محور العملية الإرشادية إذ توجه كل الجهود نحو تسجيله وإعداده بالشكل الصحيح لتلافي وقوع أي مشكله ولكنه يستأثر بدور مهم في هذا الموضوع.
كل العملية التعليمية وما يتبعها من خطوات وإجراءات تبدأ من عند الطالب فإن كان إرشادا أو تسجيلاً أو حضور أًللدراسة أو أداءاً لواجباته الأكاديمية يجب على الطالب أن يبادر إلى الشروع بالإجراءات اللازمة ، وعندما تعترضه أي مشكله عليه مراجعة جهة الاختصاص حسب نوع المشكلة.

٤.٤المرشدالأكاديمي

المرشد الأكاديمي يقصد به عضو هيئة التدريس المكلف بإرشاد عدد معين من الطلاب في كل ما يتعلق بشؤونهم الأكاديمية والإجتماعية والصحية منذ قبولهم بالكلية .

أوهوعضو هيئة التدريس الذي يقوم بتقديم العون للطالب في أثناء عملية التسجيل كما يتولى مساعدة الطالب وإرشاده في تنفيذ خطته الدراسية وتقديم النصح له في الأمور المتعلقة بالشؤون الأكاديمية ومتابعة تحصيله العلمي وإختيار المساقات التي سوف يسجلها ويساعده في حل ما قد يعترضه من مشكلات. ويوزع الطلبة على المرشدين في القسم العلمي من قبل الكلية.

٥.٤ مهام المرشد الأكاديمي

يقوم المرشد بتخصيص زمن محدد للإرشاد يضمن في الجدول الأسبوعي ويعلن للطلاب. على المرشد أن يكون على إستعداد للإستماع بتروٍ لطلابه وعليه أن يهتم بشؤونهم الأكاديمية والإجتماعية و عليه أنيقوم بشرح نظم ولوائح الإمتحانات ودليل الإرشاد الأكاديمي وملحقاته للطالب،يتابع المرشد السير الأكاديمي للطالب منذ تسميته مرشداً له حتى تخرجه،يراجع المرشد الآداء المستمر للطالب وعمله ليكون على دراية بمدى إنتظام الطالب في الدراسة وفي آدائه للواجبات الأكاديمية التي يكلف بها،يقوم المرشد بإعداد ملف خاص لكل طالب يحتفظ فيه بآدائه الأكاديمي وسجلات درجاته في المواد وأي أوراق أخرى تتعلق بالطالب.

يقوم المرشد عند بداية كل فصل دراسي بالتوقيع على إستمارة تسجيل الطالب،يقوم المرشد بنصح الطالب وتوجيهه عند تحديد نوع القيد وبالتوقيع على الإقرار عند الحاجة له وتبصير الطالب بكافة العواقب، يقوم المرشد بمساعدة الطالب على تنسيق جدول امتحاناته متى ما إستدعى الوضع الاكاديمي للطالب ذلك ليتوافق مع جدول الامتحانات المعلن، ويقوم المرشد بدراسة أي مشكله لدى طلابه ويسعى لمعالجتها بالتنسيق مع القسم المعين.

٦.٤المشاكل التي تواجه المرشد الاكاديمي

- عدم توفر الوقت الكافي لدى المرشد للقيام بالعملية الإرشادية.
- أحياناً يتعذر على المرشد الأكاديمي القيام بمهة الإرشاد بسبب كثرة المهام المسندة إليه.
- يشكو بعض المرشدين من زيادة عدد طلابهم.
- عدم توفر التدريب الكافي للمرشدين الأكاديميين والإعتماد على قدراتهم الشخصية في إكتساب الخبرة في الإرشاد.

٧.٤ بعض السمات الأساسية للساعات المعتمدة

ينجح الطالب في المقرر إذا حصل على 42% درجات المقرر،ويعتبر الحصول على50% من درجات مقرر إزالة الرسوب نجاحاً بتقدير ضعيف.

يسمح الطالب بالإنتقال من فصل إلى التالي بما لايزيد عن 4 من مقررات (أو 10من ساعات) الرسوب والبديل ويجوز للكلية السماح بالإنتقال مع تأجيل الجلوس للإمتحان في ما لا يزيد عن 4 من هذه المقررات (أو 10من ساعات) شريطة ان لا تزيد ساعات الرسوب عن المؤجلة 4 ساعات معتمده.

٨.٤ حساب النتيجة:

يحسب المعدل الفصلي عن طريق الحصول على الدرجات التي حصل عليها الطالب في نهاية الفصل مقسومة على 25 لحساب الدرجة النقطية لكل مقرر (وفي حالة إزالة الرسوب يتم القسمة على 30) ثم تحسب نقاط المقرر بضرب الدرجة النقطية في عدد الساعات المعتمدة للمادة ثم نقوم بقسمة مجموع نقاط المقررات على مجموع الساعات المعتمدة الفصلية لحساب المعدل الفصلي.

المعدل الفصلي=النقاط الفصلية÷الساعات المعتمدة الفصلية

(مقرباً إلى منزلتين عشريتين)

أما المعدل التراكمي فيتم حساب النقاط الفصل الدراسي الحالي زائدا النقاط التراكمية والساعات المعتمدة زائداً الساعات التراكمية وقسمتهما أما في حالة وجود رسوب يتم حساب نقاط المقرر للفصل زائداً نقاط المقرر لمواد الرسوب لحساب النقاط الفصلية وكذلك الساعات الفصلية عبارة عن مجموع الساعات المعتمدة للفصل زايداً مجموع الساعات المعتمدة لإزالة الرسوب ثم يحسب المعدل التراكمي

المعدل التراكمي=النقاط التراكمية÷الساعات المعتمدة التراكمية

(مقرباً إلى منزلتين عشريتين)

الباب الخامس

جمع البيانات

٥.١مقدمه

في هذا البحث تم التعامل مع إستمارة التسجيل للطالب الأكاديمي وتم التعامل مع البيانات المتعلقة بالمواد التي لم ينجح بها مع المعدل التراكمي (GPA) والوضع الأكاديمي.

٥.٢ اختيار البيانات

تم الحصول على البيانات من مكتب المسجل للأقسام المختلفة في كلية علوم الحاسوب جامعة السودان، وتتمثل هذه البيانات في إستمارات التسجيل للطلاب خلال دراستهم الجامعية علي مدار الأربع سنوات الشكل رقم(٤-١)يوضح إستمارة الطالب الأكاديمي، التي تحتوي على ثلاث أجزاء الجزء الأول خاص بالطالب ويحتوي على إسم الطالب ورقم التسجيل والقسم الذي ينتمي إليه الطالب و الفصل الدراسي الذي درسه. الجزء الثاني يحتوي على عدد مقررات البديل والرسوب التي إمتحنها الطالب يحددها المرشد الأكاديمي مع الطالب ويحدد الفصل الذي سيمتحنها فيه، ويحتوي إسم المرشد وتوقيعه وإعتماد رئيس القسم،أما الجز الثالث خاص بمسجل الكلية.

تهمنا فقط الإستمارات التي تم التعليق عليها من قبل المشرفين الأكاديميين لوضع خطة دراسية للطالب تتناسب مع معدله والمواد التي واجه مشاكل بها أما الطلاب الذين تم التوقيع علي إستماراتهم من دون تعليق أو ملاحظات ليسجلوا للفصل الدراسي الجديد بنفس الخطة الدراسية الأصلية للكلية لا يحتاجون للإرشاد الأكاديمي لوضع خطة دراسية أخري. والإستمارات يمكن أن تكون متتالية أو متفرقة علي مدار الفصول الدراسية للطالب.

٥.٣ تهيئة البيانات لقواعد الإرتباط

بعد إكتمال جمع البيانات من مصادرها وإستبعاد الإستمارات للطلاب الناجحين من مجموعة الإستمارات التي كانت في شكل ملفات ورقيه قمنا بحوسبتها في منظومة (Excel) الصف الأول عبارة عن المقررات الدراسية التي درسها الطالب منذ إلتحاقه بالكلية إلي التخرج،الشكل رقم(٥-٢)يوضح نموذج للبيانات في (Excel) من أول فصل دراسي إلي أخر فصل.أما العمود الأول يمثل أرقام الطلاب الجامعية والعمود الثاني يمثل المعدلات التراكمية للطلاب علي حسب الفصل الذي درسوه، بقية الخلايا في (Excel)تم وضع الرقم واحد في الخلية لتمثل المادة التي درسها الطالب المعين علي حسب الفصل الدراسي والخلايا الخالية تعني أن المواد المعينة في الفصل المعين لن يدرسها الطالب إما لأنه درسها ونجح فيها أو أنها تدرس في فصل قادم. قمنا بعمل صورة لملف (Excel)من دون إظهار أرقام الطلاب الجامعية بسبب سرية هذه البيانات. ثم تمت تهيئة الملف لتناسب أداة تنقيب البيانات (weka).

استمارة تسجيل طالب نظامي (نظام الساعات المعتمدة)

العام الدراسي .....

الجزء الأول: بيانات خاصة بالطالب:

(١)الاسم: .....(٢) رقم التسجيل:.....

(٣)القسم :..... (٤) الفصل الدراسي:.....

(٥) السنة الدراسية : .....

الجزء الثاني خاص بالقسم:

(١)النتيجة السابقة:.....

(٢) عدد مقررات البديل والرسوب (وأسماؤها) التي سوف يجلس الطالب فيها (يحددها الطالب مع المرشد الأكاديمي):

أ/ مع امتحانات الفصل: .....السنة: .....(لا تزيد عن أربعة ولا تزيد ساعاتها عن ١٠ ساعة معتمدة):

م	المقرر	عدد الساعات	م	المقرر	عدد الساعات
١			٣		
٢			٤		

ب/ مع امتحانات الفصل: ..... السنة: ..... (مقررات مؤجلة لا تزيد عن أربعة ولا تزيد ساعاتها عن ١٠ ساعات معتمدة):

م	المقرر	عدد الساعات	م	المقرر	عدد الساعات
١			٣		
٢			٤		

(٣) اسم المرشد الأكاديمي:.....التوقيع:..... التاريخ:.....

(٤) اعتماد رئيس القسم..... التاريخ:.....

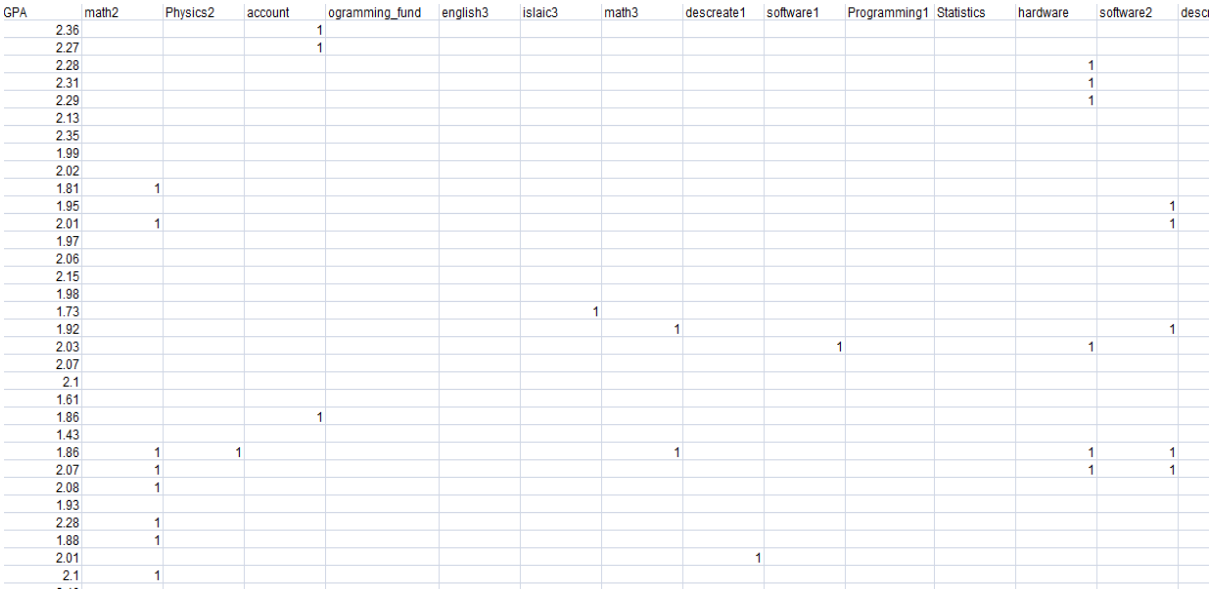
الجزء الثالث:- خاص بمسجل الكلية

أ/ هل استوفى الطالب شروط التسجيل؟:.....

ب/ هل وقع الطالب الذي لديه إنذار ثان على الإقرار؟.....

توقيع المسجل :..... التاريخ: .....

الشكل (٥-١) لاستمارة الطالب الأكاديمي

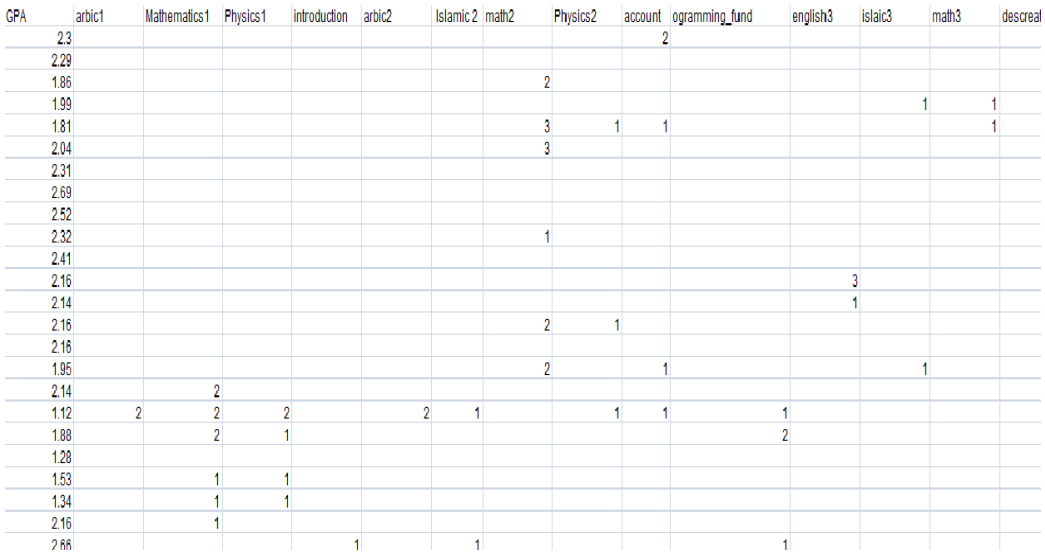


GPA	math2	Physics2	account	ogramming_fund	english3	islaic3	math3	descreate1	software1	Programming1	Statistics	hardware	software2	desc
2.36			1											
2.27			1											
2.28												1		
2.31												1		
2.29												1		
2.13														
2.35														
1.99														
2.02														
1.81	1													
1.95													1	
2.01	1												1	
1.97														
2.06														
2.15														
1.98														
1.73						1								
1.92							1						1	
2.03									1			1		
2.07														
2.1														
1.61														
1.86			1											
1.43														
1.86	1	1					1					1	1	
2.07	1											1	1	
2.08	1													
1.93														
2.28	1													
1.88	1													
2.01								1						
2.1	1													

الشكل (٥-٢) نموذج لبيانات (Apriori)

٤.٥ تهيئة البيانات للفلترة التعاونية

العمود الأول يمثل أرقام الطلاب الجامعية والعمود الثاني يمثل متوسط المعدلات التراكمية للطلاب، أما بقية الخلايا في (Excel)تم وضع الأرقام المختلفة لتمثل عدد المرات للمواد التي جلس الطالب لتقديمها. الشكل رقم(٥-٣)يوضح نموذج للبيانات في (Excel) للفلترة التعاونية.



GPA	arbic1	Mathematics1	Physics1	introduction	arbic2	Islamic 2	math2	Physics2	account	ogramming_fund	english3	islaic3	math3	descreat
2.3									2					
2.29														
1.86							2							
1.99												1	1	
1.81							3	1	1				1	
2.04							3							
2.31														
2.69														
2.52														
2.32							1							
2.41														
2.16												3		
2.14												1		
2.16							2	1						
2.16														
1.95							2		1				1	
2.14			2											
1.12	2		2	2		2	1		1	1	1			
1.88			2	1							2			
1.28														
1.53			1	1										
1.34			1	1										
2.16			1											
2.66				1		1				1				

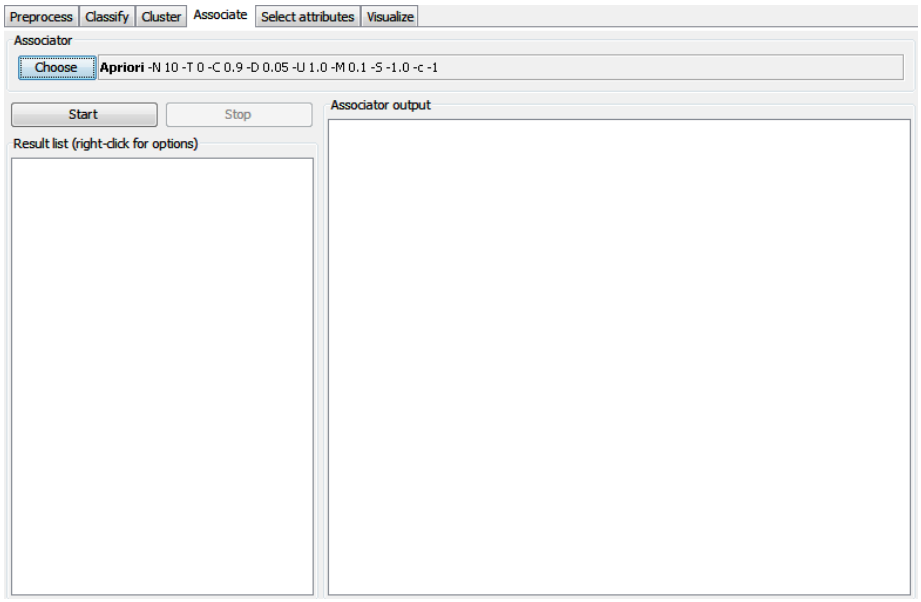
الشكل (٥-٣) نموذج لبيانات (K-NN)

٥.٥ تنقيب البيانات

وتم في هذه المرحلة إستخدام تقنية الفلتره التعاونية (collaborative filtering) بإستخدام خوارزمية (KNN)وأيضاً إستخدام قواعد الإرتباط بخوارزمية (Apriori).

٥.٦طريقة إختيار خوارزمية (Apriori)

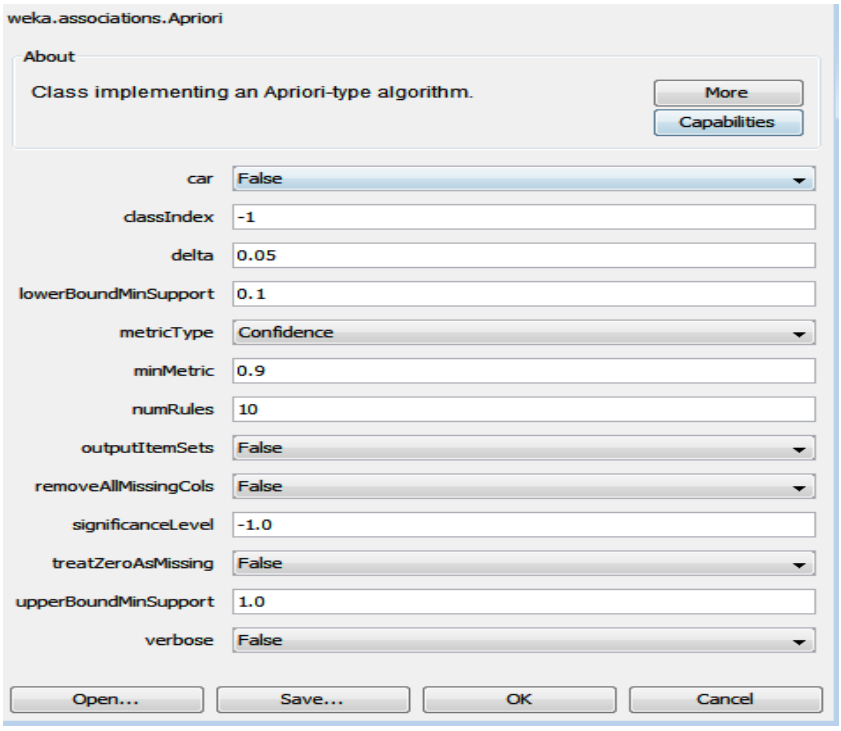
بعد أن يتم تحميل البيانات نقوم بتحديد الصفات(attributes) التي نريد إجراء عملية التجميع عليها ويمكن إختيار جزء منها أو إختيار جميعها وكذلك الحذف.ثم الرجوع لأعلي القوائم لإختيار (Associate) وبعدها إختيار (start)لتبدأ خوارزمية (Apriori)بالعمل وإظهار المخرجات والشكل(٥-٣) يوضح طريقة إختيار الخوارزمية.



الشكل(٥.٤) طريقة إختيار خوارزمية(Apriori)

طريقة إختيار (support) و(confidence)

في خوارزمية (Apriori) أجريت تجارب عديدة علي البيانات وأعطت نتائج مختلفة وإختلاف النتائج يعتمد علي تغيير (support) و(confidence) وتغيير الصفات، والمشكلة الأساسية في الخوارزمية هي إختيار (support) و(confidence) المناسب بسبب ذلك النتائج التي تم التحصل عليها منها ماهو منطقي ومنها ماهو غير منطقي والشكل(٥-٤) يوضح طريقة إختيار (support) و(confidence).



الشكل(٥.٥) شاشة إختيار(support) و(confidence)

وبعد الاطلاع علي النتائج قمنا بتحديد القواعد التي يظهر فيها المعدل التراكمي(GPA) مع المواد التي جلس لامتحانها مما يساعد في تحديد أي مواد يمكن الجمع بينها وأي المواد التي اذا تم الجمع بينها يؤثر ذلك بشكل سلبي علي نتيجة الطالب، ويمكن معرفة ذلك من خلال المعدل التراكمي لأنه يوضح نتيجة المواد التي جلس لامتحانها.

الشكل (٥-٥)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 مع تثبيت (Support) وتغيير(confidence):

Confidence	Support	n.Rules	Large item set	n.item sets
0.1	0.001	14	4	4
0. 2	0.001	2	2	2
0. 3	0.001	1	2	2

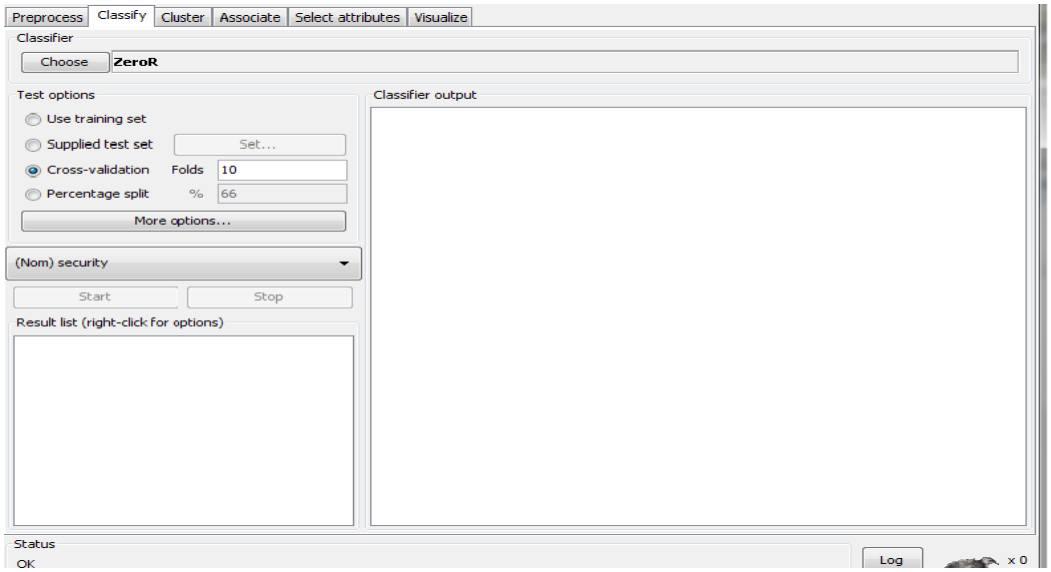
الشكل (٥-٥) يوضح النتائج المتحصل عليها

ومن النتائج لاحظنا كلما زادت نسبة (Confidence) قل عدد (item set) وكلما قل (item set) كانت العلاقات بينهم أكثر منطقية،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.

٧.٥ كيفية عمل خوارزمية (k-NN)

بعد أن يتم تحميل البيانات لعمل خوارزمية(K-NN) نقوم بتحديد الصفات(attributes) التي نريد إجراء عملية التجميع عليها، إختيار (choose) في (classify)لتحديد الخوارزمية بالضغط علي (lazy) وبعدها (IBK)وهي خوارزمية الفلترة التعاونية أو الجار الأقرب وأخيراً إختيار العناصر المراد إختبارها أسفل (more options) ثم إختيار (start) لتبدأ الخوارزمية بالعمل وإظهار المخرجات كما في الشكل (٥-٦) .

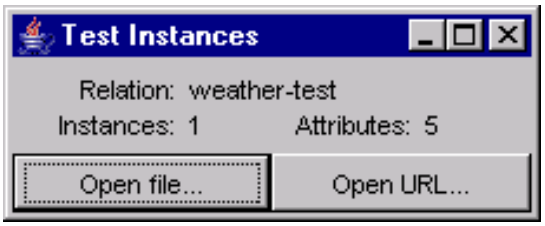


الشكل (٥-٦)شاشة(classify)

تحديد خيارات الاختبار(test options)

١. إستخدام مجموعة التدريب(Use training set): يعني إستخدام مجموعة التدريب (الملف الذي قمت بتحميله في (Preprocess) للإختبار.

٢. التزود بمجموعة إختبار (Supplied test set): يعني أنه يمكنك تحديد ملف مع بيانات الإختبار. للقيام بذلك تحديد الخيار والنقر على تعيين.وتظهر نافذة صغيرة تسمى حالات الإختبار التي تسمح لك لتحميل ملف إختبار ثم يظهر لك إسم العلاقة وعدد من الصفات والحالات. أنقر على الملف المفتوح (open file) وتحميل ملف إختبار. يمكنك الحصول على ما يلي:



٣. عبر التحقق من الصحة (Cross-validation) :هو أسلوب تنبؤ للتحقق من صحة نموذج لتقييم كيف يمكن للنتائج أن تحلل الإحصائيات والتعميم إلى مجموعة بيانات مستقلة.

٤. نسبة الانقسام (Percentage split) :تعني أنه سيتم تقييم نتائج التصنيف على مجموعة الإختبار التي هي جزء من البيانات الأصلية. الإنقسام الإفتراضي (كما هو موضح في منطقة النص بجانب الخيار) هو ٦٦٪، وهو ما يعني أن ٦٦٪ من البيانات تذهب للتدريب و ٣٤٪ للإختبار.

الباب السادس

مناقشة النتائج والتوصيات

١.٦مقدمة

في هذا الباب سوف نناقش نتائج كلاً من خوارزمية(Apriori) و(k-nn)

٢.٦نتائج خوارزمية (Apriori)

```
31. BS 126=1 CS 215=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
32. BS 126=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
33. GPA=1.37 BS 126=1 CS 215=1 MAT 222=1 1 ==> CS 225=1 1    conf:(1)
34. GPA=1.37 BS 126=1 CS 215=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
35. GPA=1.37 BS 126=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
36. GPA=1.37 CS 215=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
37. GPA=2.48 CS 215=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
38. BS 126=1 CS 215=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
39. GPA=1.37 BS 126=1 CS 215=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1    conf:(1)
40. CS 215=1 MAT 222=1 4 ==> CS 225=1 3    conf:(0.75)
41. CS 215=1 EN 223=1 3 ==> CS 225=1 2    conf:(0.67)
42. MAT 222=1 EN 223=1 3 ==> CS 225=1 2    conf:(0.67)
43. MAT 222=1 6 ==> CS 225=1 3    conf:(0.5)
44. GPA=1.37 2 ==> CS 225=1 1    conf:(0.5)
45. GPA=2.01 2 ==> CS 225=1 1    conf:(0.5)
46. GPA=2.44 2 ==> CS 225=1 1    conf:(0.5)
47. GPA=2.48 2 ==> CS 225=1 1    conf:(0.5)
48. GPA=2.07 CS 215=1 2 ==> CS 225=1 1    conf:(0.5)
49. GPA=2.48 EN 223=1 2 ==> CS 225=1 1    conf:(0.5)
50. GPA=1.98 3 ==> CS 225=1 1    conf:(0.33)
51. CSS 221=1 3 ==> CS 225=1 1    conf:(0.33)
52. CS 215=1 14 ==> CS 225=1 4    conf:(0.29)
53. EN 223=1 20 ==> CS 225=1 5    conf:(0.25)
54. GPA=2.07 4 ==> CS 225=1 1    conf:(0.25)
55. GPA=2.23 4 ==> CS 225=1 1    conf:(0.25)
56. GPA=0 5 ==> CS 225=1 1    conf:(0.2)
57. GPA=2.14 5 ==> CS 225=1 1    conf:(0.2)
58. ENG 212=1 5 ==> CS 225=1 1    conf:(0.2)
59. BS 126=1 8 ==> CS 225=1 1    conf:(0.13)
```

الشكل(١.٦) نتائج خوارزمية(Apriori)

الشكل(١.٦) يوضح جميع نتائج خوارزمية (Apriori) وقد قمنا باختيار بعض القواعد لمناقشتها ومن القواعد المعتمدة.

١. GPA=1.37 BS 126=1 CS 215=1 MAT 222=1 1 ==> CS 225=1 1   conf:(1)

2. GPA=1.37 BS 126=1 CS 215=1 EN 223=1 1 ==> CS 225=1 1   conf:(1)

3. GPA=1.37 BS 126=1 CS 215=1 MAT 222=1 EN 223=1 1 ==> CS 225=1 1

١.٢.٦مناقشة نتائج خوارزمية Apriori

١.الطلاب الذين يرسبون في مادة المحاسبة ومادة الإحصاء سوف يرسبون أيضا في مادة الهياكل المتقطعة ومعدلهم سوف يكون أقل من 1.40.

٢.الطلاب الذين يرسبون في مادة الإحصاء ومادة المكونات المادية سوف يرسبون أيضا في مادة الهياكل المتقطعة ومعدلهم سوف يكون أقل من 1.40.

٣.الطلاب الذين يرسبون في المواد الإحصاء والهياكل المتقطعة١ و المكونات المادية سوف يرسبون أيضا في مادة الهياكل المتقطعة ٢ ومعدلهم سوف يكون أقل من 1.40.

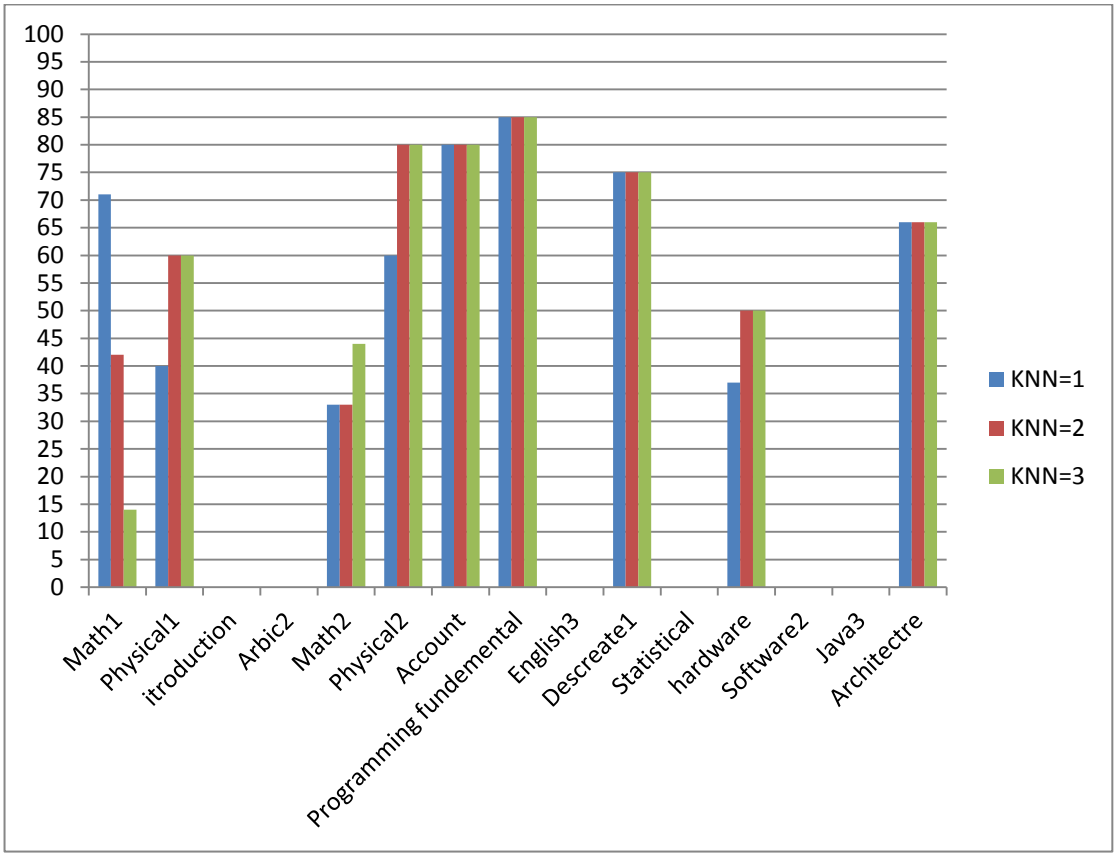
٢.٢.٦التوصيات

١. إذا رسب الطالب في مادتي المحاسبة والإحصاء لا ينصح بالجلوس لهذه المواد معا بل يجب أن يمتحن واحدة فقط حتى لا يرسب المادتين ويرسب في مادة ثالثة.

٢.إذا رسب الطالب في مادتي الإحصاء والمكونات المادية لا ينصح بالجلوس لهذه المواد معا بل يجب أن يمتحن واحدة فقط حتى لا يرسب المادتين ويرسب في مادة ثالثة.

٣.إذا رسب الطالب في المواد الإحصاء والهياكل المتقطعة١ والمكونات المادية لا ينصح بالجلوس لهذه المواد بل ينصح بأن يعيد السنة الدراسية حتى لا يأخذ إنذار بسبب المعدل.

٣.٦مخطط يوضح نسب خوارزمية (KNN)



الشكل(٢.٦) مخطط نسب خوارزمية(KNN)

الشكل(٢.٦) يوضح مخطط نسب خوارزمية(KNN) للطلاب الراسبين في المواد عندما تكونknn=1,2,3، K=1 هذا يعني أن الخوارزمية تقارن بمتوسط معدل تراكمي واحد فقط للطلاب الراسبين بعد تقديمهم للمواد مع بقية متوسطات المعدلات التراكمية التي هي مجاوره أو مقاربه له، أما k=1,2,3,٠٠٠ هذا يعني أنه يقارن بمتوسط معدلين أو ثلاثة أو أكثر من ذلك علي حسب K.

١.٣.٦مناقشة نتائج المخطط

١- في math1 هذا يعني أن متوسط معدل الطلاب في مادة الرياضيات١ عندما k=1 النسبة ٧١% وعندما k=2 نسبة متوسط معدل الطلاب ٤٢% وعندما k=3 نسبة متوسط معدل الطلاب ١٤% وهكذا في بقية المخطط.

٢- في بعض المواد تظهر نسب k (1,2,3)=0 وهذا يعني أنه لاتوجد علاقة بين متوسطات معدلات الطلاب الراسبين.

٣- يلاحظ بعض المواد لديها نسب k (1,2,3) ثابتة، وهذا يدل علي أن الطلاب الراسبين في هذه المواد معدلاتهم متقاربة.

٢.٣.٦التوصيات

١- للحصول علي أفضل النتائج يجب استخدام أكبر قدر من البيانات وعمل ملف كامل لكل طالب.

٢- حوسبة البيانات التي ستستخدم لهذا النظام.

٣- دمج كل من خوارزميتي (K-NN) و(Apriori) في نظام خليط للحصول على نتائج أفضل.

الخاتمة

بعد تحليل النتائج ومناقشاتها والاطلاع علي الحالات الأكاديمية للطلاب قيد الدراسة ظهرت أهمية نظم التوصية لتجنب إعطاء توصيات غير ملائمة لحالة الطلاب بسبب قرارات تخص خطتهم الدراسية والاستفادة من الحالات لطلاب اتبعوا الإرشاد الجيد.

وأثبتت النتائج أن من الصعب التخلي عن المرشد الأكاديمي ولكن يمكن مساعدة المرشدين الجدد ببعض الأمور التي تتعلق بمساعدة الطالب الأكاديمي المتعثر أكاديميا حتي يستطيع تحسين أداءه ومواصلة مشواره الدراسي.

المراجع

المراجع

[1] *تنقيب في البيانات* (2015, 1أبريل). Retrieved from ويكيبيديا:
https://ar.wikipedia.org/wiki/

البيانات_في_تنقيب

[2]. *التنقيب في البيانات واستخراج المعرفة*.(٢٠١٥) هيام حايك. Retrieved 129, 2015, from نسيج:

http://blog.naseej.com/2014/03/27/التنقيب-عن-البيانات

[3] تطوير اليات جديدة للتنقيب في . (2009). آصف شميس.راكان رزوق،د.د. فادي خلوف، أ: المعطيات لادارة علاقات الزبائن في بيئة مصرفية. *مجلة جامعة دمشق للعلوم الهندسية*, 15.

[4]*Apriori algorithm*. (2015, September 18). Retrieved 12 9, 2015, from Wikipedia: https://en.wikipedia.org/wiki/Apriori_algorithm

[5]Michael D. Ekstrand, John T. Riedl,Joseph A. Konstan. (2011). *Collaborative Filtering Recommender Systems.* Human–Computer Interaction.

[6]*Collaborative filtering*. (2013, November 29). Retrieved 8 14, 2015, from Wikipedia: https://en.wikipedia.org/wiki/Collaborative_filtering

[7]What is R? (n.d.). Retrieved 9 21, 2015, from r: https://www.r-project.org/about.html

[8]Data Mining using R. (n.d.). Retrieved 9 22, 2015, from Rattle: http://rattle.togaware.com/RattleBrochure.pdf

[9] *RapidMiner* . (2014, October 8). Retrieved jluy 23, 2015, from Wikipedia: https://en.wikipedia.org/wiki/RapidMiner

[10]What is Apache Mahout? (n.d.). Retrieved 9 21, 2015, from mahout: http://mahout.apache.org

[11] *PSPP*. (2015, June 20). Retrieved June 12, 2015, from Wikipedia: https://en.wikipedia.org/wiki/PSPP

[12] *What is R?* (n.d.). Retrieved 9 21, 2015, from r: https://www.r-project.org/about.html

[13] Abernethy, M. (2010, May 11). *Data mining with WEKA*. Retrieved May 8, 2015, from developerworks: http://www.ibm.com/developerworks/library/os-weka2/

[14] *تنقيب في البيانات* (2015, أبريل1). Retrieved 9 6, 2015, from ويكيبيديا: https://ar.wikipedia.org/wiki/تنقيب_في_البيانات

[15] Aleksandra Kla, Boban Vesin,Mirjana Ivanovi,Zoran Budimac. (2010, November 2). E-Learning personalization based on hybrid recommendation strategy and. *Computers & Education* , 15.

[16] Sunita B. Aher, L.M.R.J. Lobo. (2013, April 25). Combination of machine learning algorithms for recommendation of. *Knowledge-Based Systems* , 14.

[17] Sanjog Ray, Anuj Sharma. *A Collaborative Filtering Based Approach for.* Madhya Pradesh.

[18] Ying Zhang, Samia Oussen,Tony Clark, Hyeonsook Kim. *USE DATA MINING TO I.* London.

[19] العلمية, م. ا. (2012). *دليل الارشاد الاكاديمي*. الخرطوم: جامعة السودان للعلوم والتكنولوجيا.