هل تؤثر البيانات الديموغرافية على تخرج الطالب؟ دراسة تحليلية باستخدام تقنيات تنقيب البيانات

أ. نورس كاظم يوسف
كلية الاداب والعلوم قصر الاخيار
جامعة المرقب

أسماء عبد الله المحمودي
كلية الاداب والعلوم قصر الاخيار
جامعة المرقب

الملخص:

يهدف البحث إلى تحليل واقع خريجي طلبة كلية الآداب والعلوم قصر الأخيار ودراسة أسباب تأخرهم في الدراسة باستخدام تقنيات التنقيب في البيانات الديموغرافية مع اختيار خورزميتي شجرة القرار Decision Trees والعنقدة Clustering في التحليل وذلك لاستخلاص أنماط جديدة من المعرفة تساعد متخذي القرار في تطوير العملية التعليمية. وتم جمع عينة من بيانات طلبة المرحلة المنتهية من الدراسة من مختلف أقسام الكلية، تمثلت هذه البيانات بخصائصهم الديموغرافية التي أدى تحليلها إلى الوصول إلى مجموعة من النتائج حددت بموجبها أسباب تأخره في التخرج، منها العمر عند التسجيل وتخصص الثانوية والقسم العلمي المقبول به الطالب وغيرها. ولأن البحث ركز بدرجة كبيرة على الطالب فإن مجموعة من التوصيات المرتبطة بالنتائج وضعت لحث متخذي القرار في الكلية للاهتمام بسياسة القبول والتسجيل في الكلية وبناء مستودع بيانات إلكتروني يحوي جميع بيانات الطلبة المسجلين.

الكلمات المفتاحية : تنقيب البيانات، العوامل الديموغرافية، تخرج الطالب، شجرة القرار، العنقدة.

Are Demographic Data affecting the Student's Graduation? An Analytical Study Using Data Mining Techniques

Abstract

The research aims at analyzing the reality of graduates of the Faculty of Arts and Sciences Kasr Khiar, and to study the reasons of delay in the graduation, using data mining techniques of the demographic data by the decision tree and clustering algorithms in order to derive new forms of knowledge to help decision makers in developing the educational process. Sample data was collected from final stages from different departments, these data represents the demographic attributes to be analysed to lead to a set of results that determined the reasons of the delay in graduation, such that the age, kind of secondary school and the department. Because the research focused on the student, a series of recommendations were placed to urge decision makers in the faculty to focuce on admission's policy and build an electronic data wearhouse containing data of all students enrolled.

1.المقدمة

تتطور تكنولوجيا المعلومات بشكل ملحوظ بالاضافة إلى تزايد حجم البيانات المستخدمة في المؤسسات، و تأتي الخطوة الأهم وهي كيفية استثمار هذه الكميات الضخمة من البيانات بشكل ذكي في تحليل ما مضى من معاملات وأحداث وإجراءات داخل المؤسسة، ومن ثم يمكن أن يتم رسم السياسات والخطط الإستراتيجية على بصيرة ورؤية وموروث معلوماتي يصف مواضع نجاح وإخفاق المؤسسة ويتنبأ بفرص التحسين المستقبلي.مع هذه الكميات الضخمة من البيانات فإن الطرائق التقليدية لتحليل البيانات والتي هي مزيج من الطرائق الإحصائية وبعض النظم الحاسوبية المصممة لإدارة قواعد البيانات باتت تعاني الكثير من المشكلات في التعامل مع هذا النوع من البيانات (الجناعي، 2011، 58).

حيث اتجه الباحثون لإيجاد طرق بديلة يمكن أن نسميها بالإستنباطية لتقوم بدور يحاكي دور محلل البيانات لاستكشاف العلاقات والأنماط السائدة في البيانات كما تصفها سجلات البيانات الفعلية لا كما يفترضها محلل البيانات. وكما هو متوقع تستخدم هذه الطرائق تقنيات وخوارزميات ذكية – أي أنها تحاكي نمط الاستنباط عند الانسان – فتلاحظ وتعمم وتستنبط

وأخيراً تستنتج .ومن بين العلوم التطبيقية الحديثة في هذا المجال يأتي علم اكتشاف المعرفة في قواعد البيانات Knowledge Discovery in Databases (KDD) وعلم التنقيب في البيانات Data Mining على رأس هذه العلوم في توفير أطرعامة وتقنيات وخوارزميات بل وأدوات مؤتمنة توجه وتسهل إجراء تحليلات ذكية وعميقة ومعقدة، واستكشاف أرقى أنواع المعلومات، والتي تسمى في هذا المجال بـ "المعرفة"(Knowledge) ومن ثم توفيرها لصناع القرار بسرعة قياسية وبجودة عالية (Michael, 1997,32).

وتعد عملية إدارة المؤسسات التعليمية من الصعوبات التي تواجه القائمين عليها وذلك لكبر حجمها وتشعب هيكليتها وتعدد مصادر. واذا ما تم التركيز على المشاكل التنظيمية فيما يتعلق بالطالب فان تحليل مستوى أداء الطالب خلال مدة دراسته يستلزم وجود نظام معلومات حاسوبي يبنى بطريقة علمية حديثة يزود بعملية تحليل حقيقي لبيانات الطلاب فيما يخص العملية التعليمية، وهذا التحليل يتطلب استخدام تقنيات حديثة في عملية اكتشاف المعرفة ومن هذه التقنيات تقنيات تنقيب البيانات التي يعد استخدامها في الوقت الحاضر من الضرورات وذلك لدقة نتائجها في مساعدة متخذي القرار في اتخاذ القرارت السريعة والصحيحة (Witten, 2005.122).

2. مشكلة البحث :

تُخرج الجامعات سنوياً أعداداً كبيرة من الطلبة وفي تخصصات مختلفة ، وبذلك تتوفر كميات كبيرة من البيانات لم يتم وضعها وتنسيقها في قاعدة بيانات في كثير من الكليات لكي يتم إستغلالها واستثمارها بشكل أمثل والاستفادة منها في معرفة سلوك الطالب أثناء مدة دراسته والعوامل والاسباب التي تؤدي إلى تأخره في الدراسة وارتباط ذلك بالمتغيرات الديموغرافية مثل العمر والجنس والسكن والتخصص العلمي وغيرها، وبالتالي لم يتم حصر خصائص الطالب وتحليل الاسباب التي تؤدي إلى تأخره أو عدم تأخره في الدراسة، وكذلك لم يتم دراسة مؤشرات أداء الطالب في النجاح في المواد الدراسية لمساعدة متخذي القرار في وضع سياسات كفيلة بتطوير العملية التعليمية والحد من ظاهرة الرسوب والتأخر في الدراسة.

3. أهداف البحث :

يقدم هذا البحث دراسة تحليلية في مجال التنقيب في البيانات الديموغرافية للطلاب باستخدام خورزميتي شجرة القرار Decision Trees والعنقدة Clustering وذلك

لاستخلاص أنماط جديدة من المعرفة تساعد متخذي القرار في تطوير العملية التعليمية. ويمكن تقسيم الهدف الرئيس لهذا البحث إلى الأهداف الفرعية الآتية:

1. استطلاع رأي عينة من الطلبة في المرحلة المنتهية من الدراسة.
2. تحليل البيانات باستخدام شجرة القرار والعنقدة وبعض برامج تحليل البيانات.
3. اكتشاف أنماط جديدة من المعرفة من بيانات الطلاب في الكلية.
4. دراسة مؤشرات نجاح الطالب.

4. منهجية البحث :

المنهج المتبع في البحث هو المنهج التحليلي باستخدام البرامج التالية:

1. SQL-Server add-ins datamining.
2. Excel.
3. SPSS.

تم استطلاع رأي طلبة المرحلة المنتهية في كلية الاداب والعلوم قصر الاخيار جامعة المرقب، نظام السنة والفصل والبالغ عددهم 200 طالب، وتم جمع البيانات من خلال استمارة أعدت لهذ الغرض. واستغرقت مدة كتابة البحث فصلين دراسيين (ربيع 2015 و خريف 2015)، واتبعت المنهجية الخطوات التالية:

1. جمع البيانات: وقد تم ذلك من خلال ملء استمارة بيانات.
2. المعالجة الأولية: وتختصر هذه المرحلة بعملية فرز البيانات المجمعة وعزل البيانات الناقصة.
3. بناء مستودع بيانات وذلك لتفريغ البيانات فيه.
4. اختيار التطبيق المناسب لعملية التنقيب في البيانات.
5. تصدير البيانات من مستودعها لغرض المعالجة وبناء تركيب البيانات المناسب للخوارزمية المستخدمة.
6. بناء نموذج لاختيار الخوارزمية المناسبة (العنقدة، شجرة القرار) وذلك لاكتشاف أنماط جديدة.
7. تحليل النتائج.

5. أهمية البحث :

تتأتى أهمية هذا البحث من أهمية معرفة تحليل خصائص الطالب، ومن أهمية استخدام أدوات تقنيات التنقيب في البيانات في تحليل البيانات الخاصة بالطلبة الخريجين وذلك :

1. للاهتمام بالطالب كونه الجيل الذي تبنى عليه الأمم.

2. تزويد المعنيين بادارات التعليم العالي بمعلومات تفصيلية عن الخريجين وخصائصهم وميولهم.
3. تحفيز القائمين على ادارة ملفات الطلبة من خلال استنباط أنماط جديدة في تخزين البيانات.
4. حث الادارات على تقديم النصائح والارشادات للطلبة لاختيار التخصص المناسب ومعرفة الطريق الصحيح للنجاح.

6. مفهوم التنقيب في البيانات:

يمكن القول إن (Data mining) هي عملية اكتشاف المعرفة من البيانات (قواعد البيانات) أو هي التنقيب عن البيانات (أحيانا تسمى إكتشاف المعرفة) وهي عملية تحليل البيانات من منظورات مختلفة واستخلاص علاقات بينها وتلخيصها إلى معلومات مفيدة. وهي عملية البحث في قواعد البيانات (مستودعات البيانات) عن معرفة غير مكتشفة وغير متوقعة أي الحصول على معرفة جديدة غير موجودة في قواعد البيانات الأصلية وتكون هذه المعرفة مهمة بحيث تساعد في اتخاذ القرار (Tom, 1997,154) (Michael,2001,50).

فالتنقيب في البيانات منهجية تجمع بين نتائج الأبحاث في الذكاء الاصطناعي، الفهم الآلي في التعرف على الأشكال، قواعد المعلومات، الرياضيات الإحصائية، واجهات الاستعمال واللغة. وهناك خوارزميات تستخدم لهذا الغرض منها شجرة القرار والعنقدة والشبكات العصبية وقواعد الارتباط وغيرها.

7. مراحل استكشاف المعرفة:

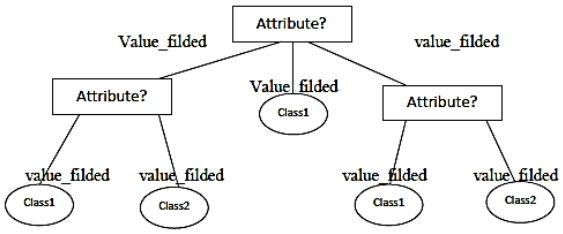
- **اختيار البيانات** Data Selection : وهي مرحلة اختيار البيانات المرشحة والملائمة للدارسة من مستودع البيانات الكلي بحسب الغرض من الدراسة.
- **تنقية البيانات** Data Preprocessing : وهي مرحلة تصفية وتجهيزالبيانات التي تحتوي على أي تشويش أو شوائب Noise من مجموعة البيانات .
- **تحويل البيانات** Data Transformation **:** وهي عملية تحويل البيانات التي تم اختيارها إلى شكل ملائم للخوارزميات والتطبيقات التي ستستخدم في الدراسة. حيث قد تشترط بعض الخوارزميات وجود البيانات على هيئة معينة قبل تطبيقها.

- التنقيب في البيانات Data Mining **:** في هذه المرحلة يتم تطبيق طرق وخوارزميات ذكية وبتسلسل مناسب لاستكشاف أنماط مفيدة.

- تقييم الأنماط Pattern Evaluation**:** بعد استكشاف الأنماط المهمة والتي تمثل المعرفة يتم تقييمها بطرق كمية ووصفية مختلفة وملائمة للتقنيات التي طبقت وباستخدام مقاييس محددة في بيئة المشكلة (الدوري، 2007، 42)(Fayyad, 1996).

8. شجرة القرارDecision Tree :

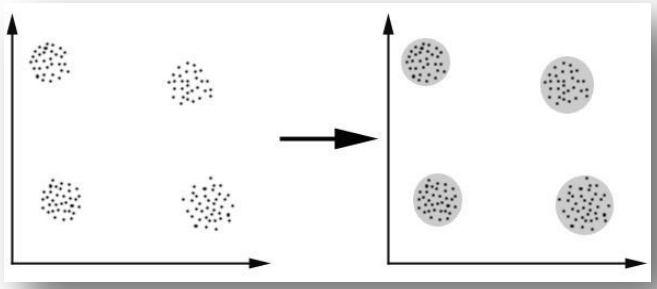
شجرة القرار هي طريقة بيانية تستخدم كثيراً لدعم عملية اتخاذ القرار في حالة عدم التأكد مع وجود احتمالات ، فهذه الطريقة تعتمد على رسم القرارات والحالات المتوقعة على شكل شجرة. وتتفرع الشجرة إلى أفرع بناءً على اختيار نختاره أو على أحداث مستقبلية لا ندري أيها يقع ، وتستخدم أشجار القرار عادة في بحوث العمليات ، وتحديداً في تحليل القرار للمساعدة في تحديد إستراتيجية على الأرجح للوصول إلى الهدف (السماني، 2004،122)(شبكة الانترنت، ويكيبيديا).



شكل (1) يوضح شجرة القرار

9. العنقدة Clustering:

هي عملية استخلاص او ايجاد مجموعة من الأشياء (الكائنات) المتشابهة فيما بينها بشكل مجاميع أوعناقيد، بحيث تكون عناصر كل مجموعة متشابهة مع بعضها بصورة أكبر ، وهذه الأشياء تسمى خصائص وتكون متشابهة فيما بينها في العنقود الواحد ومختلفة عن العنقود الآخر. وكلما زادت قوة تشابه الخصائص في العنقود الواحد تؤدي إلى أفضل عملية فصل لهذه الخصائص (Web site, Clustering).



شكل (2) يوضح عملية العنقدة

10. مؤشرات الأداء في النجاح (شبكة الانترنت، جامعة أم القرى)**:**

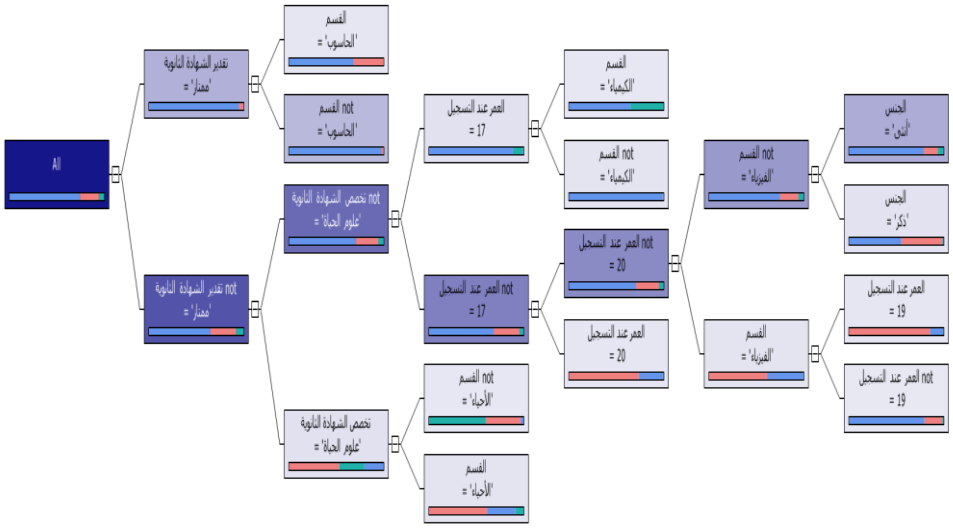
يعدّ تحسين نجاح الطلاب في يومنا هذا - سواء من خلال معدلات التخرج، أو فترة التخرج، من الأولويات للعديد من المؤسسات التعليمية، ومن هذا المنطلق نحاول في هذا البحث تحديد أهم العوامل المؤثرة على مستوى أداء الطالب في الدراسة ومدى تأثيرها من ناحية تأخر أو عدم تأخرالطالب في التخرج في الفترة المحددة. وهذه العوامل ونسميها هنا بـ (مؤشرات الأداء) وهى :

- **الرغبة** : عندما ترغب في شيء لدرجة أنك تستعد لإتمامه مهما تطلب منك الأمر ثق أنك سوف تنجح في الوصول له . فالرغبة الحقيقية هي أول خطوة فعلية في تحقيق هدفك .
- **الصبر** : قوة التحمل والإصرار على تجاوز كل الصعوبات والتحديات وعدم الإستعجال، ومن صبر ظفر.
- **نسبة الانجاز في تحضير المادة** : حقق عملا يرضيك يوميا و يقربك من هدفك، وعليك ألا تقدم على ذلك العمل إلا إذا وجدته يقربك من هدفك
- **التركيز** : عندما لا تركز على هدفك تجذبك التوافه، لذلك فإن سر نجاح العظماء هو التركيز ثم التركيز ثم التركيز.
- **الحضور والانضباط** :يعني إطاعة التعليمات والتقيد بالعمل، فالناجحون منضبطون.
- **الاهتمام بالمادة** :من حيث التنظيم وعدد ساعات الدراسة وتعدد المناهج وغيرها.
- **القراءة الجماعية** : العمل الجماعي يساعد في توليد أفكار جديدة.

- **التنظيم في الوقت** :الإلتزام بجدول زمني للمساعدة في تنظيم وأداءالمهام، ليحسن استغلال الوقت.
- **ايقاف القيد لمدة معينة.**
- **التأخر بالعمر.**

11. التحليل باستخدام خوارزمية شجرة القرار Decision Trees

في طريقة عمل خوارزمية الشجرة يكون عدد القرارات الناتجة منها مساوياً لعدد العقد النهائية في الشجرة (الأوراق) وهي العقد التي ليس لها تفرعات. وبعد تطبيق الخوارزمية والتي نتج عنها أحدى عشرعقدة منتهية وبالتالي ينتج من ذلك أحد عشرقراراً يتم توضيحها من خلال عرض البيانات والشروط اللازمة والخاصة بكل قرار وكما يلي:



شكل (3) يوضح بناء شجرة القرار لبيانات الطلاب

القرار 1:-

IF (القسم =الحاسوب) AND (تقدير الشهادة الثانوية = ممتاز) THEN (عدد سنوات التأخير=0)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
ذكر	20	الحاسوب	الأساسية	ممتاز	1
أنثى	19	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	اللغة العربية	ممتاز	1
ذكر	20	الحاسوب	الأساسية	ممتاز	1
أنثى	19	الحاسوب	الأساسية	ممتاز	0
أنثى	18	الحاسوب	الأساسية	ممتاز	0

يوضح الجدول أن أغلبية الطلبة الذين كانوا يدرسون في قسم الحاسوب والذين كانت تقديراتهم في الشهادة الثانوية (ممتاز) لم يتأخرو في الدراسة.

القرار 2:-

If (القسم= not الحاسوب) and (تقدير الشهادة الثانوية= ممتاز) Then (عدد سنوات التاخير= 0)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	19	الرياضيات	الأساسية	ممتاز	0
ذكر	18	اللغة العربية	اللغة العربية	ممتاز	0
أنثى	18	اللغة العربية	اللغة العربية	ممتاز	0
أنثى	18	اللغة العربية	اللغة العربية	ممتاز	0
أنثى	17	اللغة العربية	اللغة العربية	ممتاز	0
أنثى	19	اللغة العربية	اللغة العربية	ممتاز	0

أنثى	19	الفيزياء	الأساسية	ممتاز	0
أنثى	19	اللغة الإنجليزية	اللغة الإنجليزية	ممتاز	0
أنثى	18	التربية وعلم النفس	اللغة العربية	ممتاز	0
أنثى	18	التربية وعلم النفس	اللغة العربية	ممتاز	0
أنثى	18	الفيزياء	الأساسية	ممتاز	0
أنثى	19	الكيمياء	الأساسية	ممتاز	0
أنثى	19	الأحياء	علوم الحياة	ممتاز	0

نلاحظ في هذا الجدول أن أغلب الطلبة الذين لم يدرسوا في قسم الحاسوب و إنما درسوا في باقي الأقسام والذين كانت تقديراتهم في الشهادة الثانوية (ممتاز) لم يتأخرو في الدراسة.

القرار3:-

If (الاحياء not =القسم) and (علوم الحياة= تخصص الشهادة الثانوية) and (ممتاز =not تقدير الشهادة) Then (الثانوية (سنتان فما فوق= عدد سنوات التاخير)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
ذكر	17	الحاسوب	علوم الحياة	جيدجداً	-2
أنثى	18	الحاسوب	علوم الحياة	جيد	-2
ذكر	20	الحاسوب	علوم الحياة	جيدجداً	1
أنثى	20	الكيمياء	علوم الحياة	جيدجداً	-2
أنثى	20	الحاسوب	علوم الحياة	جيدجداً	-2
ذكر	17	الحاسوب	علوم الحياة	جيدجداً	-2
أنثى	18	الحاسوب	علوم الحياة	جيد	-2
ذكر	20	الحاسوب	علوم الحياة	جيدجداً	1
أنثى	20	الكيمياء	علوم الحياة	جيدجداً	-2

أنثى	20	الحاسوب	علوم الحياة	جيدجداً	1
ذكر	17	الحاسوب	علوم الحياة	جيدجداً	-2
أنثى	18	الحاسوب	علوم الحياة	جيد	-2
ذكر	20	الحاسوب	علوم الحياة	جيدجداً	1
أنثى	20	الكيمياء	علوم الحياة	جيدجداً	-2
أنثى	20	الحاسوب	علوم الحياة	جيدجداً	1

من الجدول نلاحظ أن أغلب الطلبة الذين لايدرسون في قسم الأحياء، والذين كان تخصصهم في الشهادة الثانوية هو علوم الحياة ، والذين لم تكن تقديراتهم في الشهادة الثانوية (ممتاز) أنهم يتأخرون في الدراسة لمدة سنتين أوأكثر.

القرار4:-

If (القسم= الاحياء) and (تخصص الشهادة الثانوية= علوم الحياة) and (تقدير الشهادة =not ممتاز) Then (الثانوية (عدد سنوات التاخير=1)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	18	الأحياء	علوم الحياة	جيدجداً	0
أنثى	18	الأحياء	علوم الحياة	جيد	1
أنثى	22	الأحياء	علوم الحياة	جيد	1
أنثى	18	الأحياء	علوم الحياة	جيدجداً	0
أنثى	21	الأحياء	علوم الحياة	جيد	1
أنثى	19	الأحياء	علوم الحياة	جيدجداً	0
ذكر	18	الأحياء	علوم الحياة	جيدجداً	-2
أنثى	19	الأحياء	علوم الحياة	جيد	1
أنثى	19	الأحياء	علوم الحياة	جيدجداً	1
أنثى	18	الأحياء	علوم الحياة	جيد	1
أنثى	20	الأحياء	علوم الحياة	جيدجداً	1

أنثى	22	الأحياء	علوم الحياة	جيدجداً	1
أنثى	18	الأحياء	علوم الحياة	جيدجداً	0
أنثى	18	الأحياء	علوم الحياة	جيد	1

من الجدول نلاحظ أن أغلب طلبة قسم الأحياء ،الذين كان تخصصهم في الشهادة الثانوية (علوم الحياة)، والذين كانت تقديراتهم في الشهادة الثانوية ما بين الجيد والجيدجداً،كانت عدد سنوات تأخرهم سنة واحدة وقلة منهم لم يتأخروا في الدراسة.

القرار5 :-

If (القسم= الكيمياء) and (العمر عند التسجيل=17) and (تخصص الشهادة الثانوية=not علوم حياة)and (عدد سنوات التاخير=0) Then(ممتازnot=تقدير الشهادة الثانوية)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	17	الكيمياء	الأساسية	جيدجداً	-2
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	-2
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0
أنثى	17	الكيمياء	الأساسية	جيدجداً	0

نلاحظ أن طلبة قسم الكيمياء الذين كانت أعمارهم17 سنة وتخصصاتهم بالشهادة الثانوية لم يكن علوم الحياة وتقديراتهم فيها لم تكن ممتاز فان أغلبهم لم يتأخروا وقلة مهم تأخروا سنتين فما فوق.

القرار6:-

If (القسم= not الكيمياء) and (العمر عند التسجيل= 17) and (تخصص الشهادة =not علوم حياة) Then (تقدير الشهادة الثانوية=not ممتاز) and(الثانوية (عدد سنوات التاخير=0)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	17	اللغة العربية	اللغة العربية	جيدجداً	0
أنثى	17	الرياضيات	الأساسية	جيدجداً	0
أنثى	17	اللغة العربية	اللغة العربية	جيدجداً	0
أنثى	17	الرياضيات	الأساسية	جيد	0
أنثى	17	اللغة الإنجليزية	اللغة الإنجليزية	جيد	0
أنثى	17	اللغة العربية	اللغة العربية	جيدجداً	0
أنثى	17	الرياضيات	الأساسية	جيدجداً	0
أنثى	17	الرياضيات	الأساسية	مقبول	0
أنثى	17	الرياضيات	الأساسية	جيدجداً	0
أنثى	17	الأحياء	الأساسية	جيد	0
أنثى	17	اللغة العربية	اللغة العربية	جيدجداً	0
أنثى	17	الرياضيات	الأساسية	جيدجداً	0

نلاحظ أن جميع الطلبة الذين لم يدرسو في قسم الكيمياء والذين كانت أعمارهم عند التسجيل هي17 سنة والذين لم يكن تخصصهم بالشهادة الثانوية علوم الحياة وتقديراتهم لم تكن ممتاز فهم لا يتأخرون في الدراسة.

القرار7:-

If (20 =العمر عند التسجيل) and (تخصص الشهادة الثانوية= not علوم الحياة) and (تقدير =not ممتاز) (عدد سنوات التاخير=1) Then (الشهادة الثانوية

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	20	الفيزياء	الأساسية	جيدجداً	1
أنثى	20	التربية وعلم النفس	الأساسية	جيدجداً	1
أنثى	20	اللغة العربية	اللغة العربية	جيدجداً	1
أنثى	20	الأحياء	الأساسية	جيد	-2
أنثى	20	اللغة العربية	اللغة العربية	جيدجداً	1
أنثى	20	التربية وعلم النفس	الأساسية	جيد	1
أنثى	20	الفيزياء	الأساسية	جيد	0
أنثى	20	التربية وعلم النفس	الأساسية	جيد	0
أنثى	20	الفيزياء	الأساسية	جيدجداً	1
أنثى	20	التربية وعلم النفس	الأساسية	جيدجداً	1
أنثى	20	اللغة العربية	اللغة العربية	جيدجداً	1
أنثى	20	الأحياء	الأساسية	جيد	-2
أنثى	20	اللغة العربية	اللغة العربية	جيدجداً	1

الطلاب الذين كانت أعمارهم عند التسجيل هي 20 سنة والذين لم يدرسوا بثانوية علوم الأحياء ولم تكن تقديراتهم بالشهادة الثانوية هي ممتاز فالنسبة الأكبر منهم يتأخرون لمدة سنة واحدة ونسبة قليلة منهم لا يتأخرون في التخرج .

القرار8:-

If (الجنس= أنثى) and (القسم= not الفيزياء) and (العمر عند التسجيل=not 17) and (تخصص الشهادة الثانوية=notعلوم الحياة) and (العمر عند التسجيل=not 20) and (عدد سنوات التأخير=0) then (تقدير الثانوية= notممتاز)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	19	الرياضيات	الأساسية	جيدجداً	0
أنثى	18	التربية وعلم النفس	اللغة العربية	جيدجداً	0
أنثى	19	الرياضيات	الأساسية	جيدجداً	0
أنثى	19	الجغرافيا	اللغة العربية	جيدجداً	0
أنثى	19	الجغرافيا	اللغة العربية	جيدجداً	0
أنثى	19	التاريخ	اللغة العربية	جيد	0
أنثى	19	الرياضيات	الأساسية	جيد	0
أنثى	18	الرياضيات	الأساسية	جيد	1
أنثى	18	الرياضيات	الأساسية	جيد	-2
أنثى	18	الرياضيات	الأساسية	جيد	1
أنثى	18	الحاسوب	الأساسية	جيدجداً	0
أنثى	19	اللغة العربية	اللغة العربية	جيدجداً	1
أنثى	19	اللغة العربية	اللغة العربية	جيدجداً	1

أما الطلبة الذين كانو من الإناث ولم تكن أعمارهم عند التسجيل 17 أو 20 سنة وكانولا يدرسون في قسم الفيزياء ولم تكن تخصصاتهم بالشهادة الثانوية علوم الحياة وأيضاً لم تكن تقديراتهم ممتاز، فانهم لم يتأخروا في التخرج وكانت نسبتهم 70% وقلة منهم يتأخرون لسنة أوأكثر.

القرار9:-

If (الجنس= ذكر) and (القسم= not الفيزياء) and (العمر عند التسجيل=not 17) and (تخصص الشهادة الثانوية=notعلوم الحياة) and (العمر عند التسجيل=not 20) and (تقدير الثانوية= notممتاز) then (عدد سنوات التأخير=0 أو 1)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
ذكر	19	الحاسوب	الهندسية	جيدجداً	1
ذكر	18	اللغة العربية	اللغة العربية	جيدجداً	0
ذكر	21	التاريخ	الأساسية	جيد	0
ذكر	21	التاريخ	الأساسية	جيد	0
ذكر	18	اللغة العربية	اللغة العربية	جيد	1
ذكر	18	اللغة العربية	اللغة العربية	مقبول	1
ذكر	18	الحاسوب	الأساسية	جيد	1
ذكر	18	اللغة العربية	اللغة العربية	جيد	0
ذكر	19	اللغة العربية	اللغة العربية	جيد	0
ذكر	18	اللغة العربية	اللغة العربية	جيد	0
ذكر	18	اللغة العربية	اللغة العربية	جيدجداً	1

وأما الطلبة الذين كانوا من الذكور والذين لم تكن أعمارهم عند التسجيل 17 و20 سنة والذين لم يدرسوا في قسم الفيزياء ولم تكن تخصصاتهم بالشهادة الثانوية علوم الحياة وأيضاً لم تكن تقديراتهم ممتاز، فهم لم يتأخروا في التخرج ونسبتهم 70% ونسبة24% منهم يتأخرون سنة واحدة.

القرار 10:-

If (العمر عند التسجيل = 19) and (القسم= الفيزياء) and (العمر عند التسجيل= 17not) and (تخصص الشهادة الثانوية=not علوم الحياة) and (العمر عند التسجيل=20 not) and (تقدير الثانوية= ممتازnot) then (عدد سنوات التأخير=1)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	19	الفيزياء	الهندسية	جيدجداً	0
أنثى	19	الفيزياء	الأساسية	جيد	1
أنثى	19	الفيزياء	الأساسية	جيد	1
أنثى	19	الفيزياء	الأساسية	جيد	1
أنثى	19	الفيزياء	الهندسية	جيدجداً	1
أنثى	19	الفيزياء	الأساسية	جيدجداً	1
أنثى	19	الفيزياء	الأساسية	جيد	1
أنثى	19	الفيزياء	الأساسية	جيد	1
أنثى	19	الفيزياء	الأساسية	جيد	1

الطلبة الذين كانت أعمارهم عند التسجيل 19سنة والذين سجلوا في قسم الفيزياء ولم تكن تخصصاتهم في الثانوية علوم الحياة وأيضاً لم تكن تقديراتهم ممتاز،فإن الأغلبية منهم يتأخرون في التخرج لمدة سنة ونسبتهم 84% .

القرار 11:-

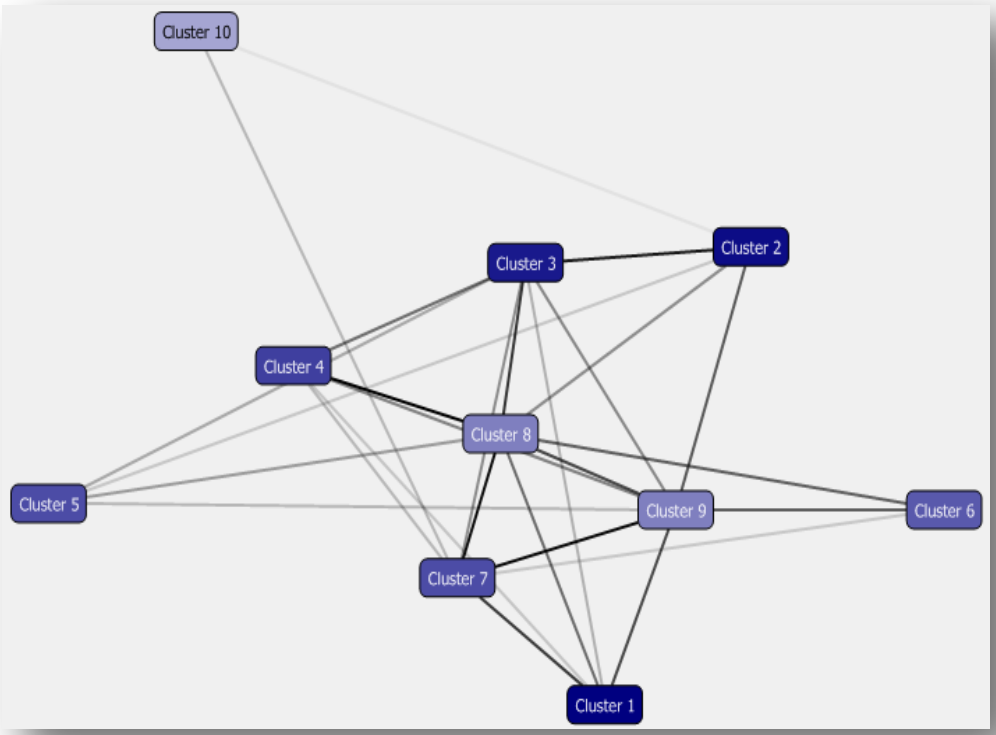
If (العمر عند التسجيل = 19not) and (القسم= الفيزياء) and (العمر عند التسجيل= 17not) and (تخصص الشهادة الثانوية=not علوم الحياة) and (العمر عند التسجيل=20 not) and (عدد سنوات التأخير=1او0) then (تقدير الثانوية= ممتازnot)

الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	عدد سنوات التأخير في التخرج
أنثى	18	الفيزياء	الأساسية	جيدجداً	1
أنثى	18	الفيزياء	الأساسية	جيد	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	1
أنثى	18	الفيزياء	الأساسية	جيد	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	1
أنثى	18	الفيزياء	الأساسية	جيد	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	0
أنثى	18	الفيزياء	الأساسية	جيدجداً	1

معظم الطلبة الذين كانت أعمارهم عند التسجيل 18سنة والذين درسوا في قسم الفيزياء ولم تكن تخصصاتهم بالشهادة الثانوية علوم الحياة وأيضاً لم تكن تقديراتهم ممتاز، هم لا يتأخرون في التخرج ونسبتهم 79% ونسبة قليلة منهم يتأخرون لسنة واحدة.

12. التحليل باستخدام خوارزمية العنقدة

بعد تطبيق خوارزمية العنقدة على بيانات الطلاب نتج عنها عشرة تجمعات (عناقيد) حيث ان كل تجمع تجمعت فيه مجموعة من الخصائص المتشابهة والتي حددت بموجبها سنوات التأخر في التخرج واتضح أن حجم العينة المختارة كان 140 من مجموع 200 وهذا الاختيار يكون عشوائيا حسب تركيب الخوارزمية وكما يلي:



شكل (4) يوضح خوارزمية العنقدة لبيانات الطلاب

جدول يوضح بيانات التجمعات العشرة المتضمنة خصائص الطالب وسنوات التاخر.

Cluster 5	Cluster 4	Cluster 2	Cluster 3	Cluster 1	Population (All)	States	Variables
13	16	17	18	19	140		Size
100 %	52 %	100 %	100 %	100 %	127	أنثى	الجنس
0 %	49 %	0 %	0 %	0 %	13	ذكر	الجنس
64 %	86 %	32 %	7 %	73 %	51	18	العمر عند التسجيل
22 %	9 %	61 %	30 %	26 %	43	19	العمر عند التسجيل
15 %	0 %	2 %	53 %	0 %	18	17	العمر عند التسجيل
0 %	0 %	0 %	5 %	1 %	11	20	العمر عند التسجيل
0 %	0 %	0 %	0 %	0 %	5	21	العمر عند التسجيل
0 %	1 %	2 %	6 %	0 %	4	16	العمر عند التسجيل
0 %	0 %	0 %	0 %	0 %	3	22	العمر عند التسجيل
0 %	4 %	0 %	0 %	0 %	3	24	العمر عند التسجيل
0 %	87 %	52 %	0 %	0 %	38	اللغة العربية	القسم
0 %	1 %	45 %	1 %	4 %	17	التربية وعلم النفس	القسم
7 %	0 %	0 %	1 %	0 %	17	الأحياء	القسم
0 %	0 %	0 %	43 %	43 %	16	الرياضيات	القسم
92 %	12 %	0 %	0 %	0 %	15	اللغة الإنجليزية	القسم
0 %	0 %	0 %	0 %	29 %	14	الحاسوب	القسم
0 %	0 %	0 %	13 %	23 %	10	الفيزياء	القسم
0 %	0 %	0 %	39 %	1 %	7	الكيمياء	القسم

Cluster 10	Cluster 9	Cluster 8	Cluster 7	Cluster 6	Population (All)	States	Variables
7	9	11	15	15	140		Size
98 %	61 %	92 %	100 %	100 %	127	أنثى	الجنس
2 %	40 %	8 %	0 %	0 %	13	ذكر	الجنس
17 %	37 %	1 %	1 %	13 %	51	18	العمر عند التسجيل
78 %	17 %	8 %	35 %	37 %	43	19	العمر عند التسجيل
0 %	0 %	1 %	43 %	6 %	18	17	العمر عند التسجيل
1 %	46 %	44 %	1 %	0 %	11	20	العمر عند التسجيل
0 %	0 %	33 %	0 %	7 %	5	21	العمر عند التسجيل
0 %	0 %	0 %	19 %	0 %	4	16	العمر عند التسجيل
0 %	0 %	0 %	0 %	22 %	3	22	العمر عند التسجيل
4 %	0 %	1 %	0 %	14 %	3	24	العمر عند التسجيل
17 %	19 %	0 %	85 %	0 %	38	اللغة العربية	القسم
15 %	0 %	55 %	2 %	0 %	17	التربية وعلم النفس	القسم
1 %	15 %	11 %	0 %	96 %	17	الأحياء	القسم
0 %	1 %	1 %	0 %	2 %	16	الرياضيات	القسم
0 %	0 %	0 %	0 %	0 %	15	اللغة الإنجليزية	القسم
45 %	57 %	0 %	0 %	0 %	14	الحاسوب	القسم
6 %	49 %	15 %	1 %	1 %	10	الفيزياء	القسم
0 %	0 %	0 %	0 %	1 %	7	الكيمياء	القسم

من الجدول السابق يمكن تلخيص التجمعات والأغلبية من الخصائص المتشابهة بناء على النسب الأكبر في الجدول التالي:

رقم التجمع	الجنس	العمر عند التسجيل	القسم	تخصص الشهادة الثانوية	تقدير الشهادة الثانوية	سنوات التأخير
1	انثى	18	الرياضيات والحاسوب والفيزياء	الاساسية	جيد جدا	0
2	اناث	19	لغة عربية وتربية وعلم نفس	لغة عربية	جيد جدا	0
3	اناث	17	الرياضيات والكيمياء	اساسية	ممتاز	0
4	اناث وذكور	18	لغة عربية	لغة عربية	جيد وممتاز	0
5	اناث	18	لغة انجليزية	لغة انجليزية	ممتاز	0
6	اناث	19	أحياء	علوم حياة	جيد وممتاز	0
7	اناث	17	لغة عربية	لغة عربية	جيد جدا	0
8	اناث	20 و 21	التربية وعلم النفس	أساسية	جيد	0
9	اناث وذكور	18 و 19	الفيزياء	أساسية	جيد	1
		18 و 20	الحاسوب	علوم حياة	جيد جدا	
10	اناث	19	الحاسوب	لغة عربية	جيد	1 و 2 فما فوق

توضح خوارزمية العنقدة عدد سنوات التأخر في التخرج حيث أن التأخر تركز على طلبة قسم الحاسوب الذين كانت تخصصاتهم في الثانوية علوم الحياة واللغة العربية وليست العلوم الاساسية. وكذلك طلبة قسم الفيزياء الذين كانت تقديراتهم في الثانوية جيد.

13. تحليل مؤشرات الأداء

بناء على البيانات التي تم الحصول عليها من الطلبة المستجوبين ، تم تصفية البيانات حسب عدد سنوات التأخير في التخرج ودراسة مؤشرات الأداء الخاصة بها ومن ثم إيجاد المتوسط الحسابي Mean لكل مؤشر علما بأنه تم جمع البيانات بناء على مقياس ثلاثي (نعم=3، نوعاما=2 ، لا=1) ومقارنة نتائج المتوسط الحسابي لكل مؤشر مع المتوسط المعياري المتمثل بالرقم =2 (نوعا ما)وكما يلي:

ر.ت	المؤشرات	المتوسط الحسابي للتأخر في التخرج		
		لا يوجد تأخر	سنة واحدة	سنتان فما فوق
-1	الرغبة في المادة	2.76	2.56	2.50
-2	الصبر	2.44	2.52	2.83
-3	نسبة الانجاز في تحضير المادة	2.60	2.40	2.42
-4	التركيز	2.76	2.68	2.25
-5	الحضور والانضباط	2.60	2.64	2.50
-6	الاهتمام بالمادة	2.76	2.56	2.67
-7	القراءة الجماعية	1.88	1.76	1.42
-8	التنظيم في الوقت	2.44	2.60	2.00
-9	ايقاف القيد لمدة معينة	1.48	1.36	1.42
-10	التأخر بالعمر	1.60	1.28	1.83

1. من الجدول أعلاه نجد أن الطلبة الذين لا يتأخرون في الدراسة ، يعتبرون أن الرغبة في المادة و التركيز فيها هما أهم المؤشرات التي تزيد من مستوى أدائهم الدراسي . بينما العوامل الأخرى مثل القراءة الجماعية والتأخر بالعمر ليس لها تأثير كبير في نظرهم.

2. أما الطلبة الذين يتأخرون لمدة سنة واحدة فيعتبرون من وجهة نظرهم أن أكثرالمؤشرات المؤثرة على دراستهم هي التركيز في المادة والتنظيم في الوقت والحضور والانضباط، بينما

يرون أن مؤشرات أخرى مثل نسبة الإنجازفي تحضير المادة أوإيقاف القيد لمدة ما قد لا تؤثركثيراً على مستوى أدائهم في الدراسة .

3. أما الطلبة الذين يتأخرون لمدة سنتين فما فوق يعتقدون أن الصبر والإهتمام بالمادة هما أبرز المؤشرات المهمة بالنسبة لهم واللتان لهما التاثيرالكبيرعلى مستوى أدائهم ،أما باقي المؤشرات كالرغبة في المادة أوالتنظيم في الوقت أوالتركيز فيعتقدون أنها أقل أهمية وأقل تأثير عليهم.

14. الاستنتاجات :

استخدمت خوارزميتان وهما شجرة القرار والعنقدة في تحليل البيانات وذلك للتأكد من صحة النتائج ولقربها من الواقع وتبين أن نتائج الخوارزميتين متقاربة إلى حد ما حيث تم دمج هذه النتائج بالاستنتاجات التالية:

1. أن النسبة الأكبرمن الطلبة الذين أكملو مرحلة الشهادة الثانوية بتقدير"ممتاز" لا يتأخرون في التخرج عن مدة أربع سنوات .
2. معظم الطلبة الذين لايتأخرون في التخرج هم من طلبة قسمي اللغة العربية واللغة الإنجليزية وأنهم كانو يدرسون في المرحلة الثانوية نفس تخصصاتهم، وكانت تقديراتهم في الثانوية بين الجيد جداً والممتاز، وأن معظمهم كانت أعمارهم 18 و19 سنة.
3. معظم الطلبة الدارسين في قسم الحاسوب يتأخرون في التخرج لمدة سنة وأكثر خاصة الذين كانت تخصصاتهم في المرحلة الثانوية هي" اللغة العربية " و" علوم الحياة ".
4. أغلبية الطلبة الدارسين في قسم الأحياء الذين يتأخرون في التخرج لسنة واحدة كانوا يدرسون في المرحلة الثانوية " العلوم الأساسية" بينماالطلبة الذين كانوا يدرسون " علوم الحياة" لا يتأخرون في التخرج.
5. كل الطلبة تقريباً الذين كانت تقديراتهم أقل من تقدير "ممتاز" والذين كانت أعمارهم17سنة وجد أنهم لا يتأخرون في التخرج، بينما الذين كانت أعمارهم 20 سنة فوجد أنهم يتأخرون في التخرج حوالي السنة.

6. معظم طلبة قسم الرياضيات الذين لا يتأخرون في التخرج كانت أعمارهم أقل من 20 سنة وكانو يدرسون في المرحلة الثانوية " العلوم الأساسية " .

7. أما بالنسبة لمؤشرات الأداء وجد أن أغلبية الطلبة الذين لا يتأخرون في التخرج يعتبرون أن درجة رغبتهم وإهتمامهم بالمادة والتركيز فيها هي العوامل الأهم والتي كانت تحدد درجة نجاحهم في المادة ، أماالطلبة الذين يتأخرون في التخرج لمدة سنة وأكثر فيعتبرون أن الصبروالإهتمام بالمادة والحضور والانضباط هي أهم العوامل المؤثرة على نجاحهم في المواد الدراسية .

8. طلبة قسم الفيزياء الذين اعمارهم 19 سنة عند التسجيل يتأخرون سنة واحدة والذين أعمارهم 18 سنة عند التسجيل قسم منهم لا يتاخرون وقسم منهم يتأخرون سنة.

9. ان علم التنقيب في البيانات يركزعلى اكتشاف أنماط جديدة من المعرفة والتي تمثلت بنتائج خوارزميتي شجرة القرار والعنقدة، يمكن لمتخذي القرار في الكلية الاستفادة منها في رسم سياسة جديدة في القبول والتسجيل تركز على البيانات الديموغرافية للطالب.

15. التوصيات :

من المعروف أن قيمة البحوث العلمية تكمن في مدى الإستفادة منها ومن نتائجها، انطلاقاً من هذا الجانب يقدم هذا البحث دراسة تطبيقية في مجال اكتشاف المعرفة للمساهمة في اعطاء رؤية واضحة لصناع القرار وتوفير قاعدة معرفية مبنية على نتائج ملموسة و قريبة من الواقع لدعم السياسات التعليمية لدى متخذي القرار في الكلية. واعتماداً على نتائج الدراسة والإطارين النظري والعملي للبحث، والملاحظات التي تم رصدها من خلال الدراسة، يتم تقديم ما يلي من توصيات :

1. الحاجة الماسة إلى بناء مستودع بيانات مترابط ومتكامل ونقي وخالٍ من الأخطاء لبيانات طلبة الكلية.
2. إعادة النظرفي سياسة القبول والتسجيل للطلبة في الكلية.
3. حث متحذي القرار في الكلية على الاهتمام بسياسة القبول والتسجيل.

المراجع والمصادر:

1. الجناعي، أواب , الحداد، الحسين ,البار، علي ,الزهاري، عمار.(2011)، استكشاف بعض الأ نماط المؤثرة في الأداء الأكاديمي لطلاب جامعة العلوم والتكنولوجيا باستخدام تقنيات التنقيب في البيانات، صنعاء، مجلة العلوم والتكنولوجيا، المجلد 16، ع3.

2. Michael J. A. Berry, Gordon Linoff , (1997).Data Mining Techniques– For Marketing, Sales, and Customer Support, Wiley, 1st Edition.

3. Witten, I.H. and Frank, E.,(2005) Data Mining: Practical machine learning tools and techniques, Morgan Kaufmann, 2nd Edition.

4. Michael J. A. Berry, Gordon Linoff, Mastering Data Mining",Wiley, (2001), 2nd Edition.

5. Tom M. Mitchell, Gordon Linoff. (1997). Machine Learning, McGraw Hill.

6. Fayyad U.,(1996). Shapiro G., and Smyth P., "From Data Mining to Knowledge Discovery indatabases.

7. الدوري، زكريا ، عبد الحسين، داليا . (2007)، تنقيب البيانات في زيادة أداء المنظمة (دراسة تحليلية في المصرف الصناعي)، بغداد، مجلة العلوم الاقتصادية والادارية، ع48 .

8. السماني، عبد المطلب أحمد ,البريهي، إياد محمد مهيوب غالب. (2004)، استخدام تقنيات تنقيب البيانات لكشف التطفل في شبكات الحاسوب، مجلة جامعة الناصر، ع4.

9. ويكيبيديا، شجرة القرار، ttps://ar.wikipedia.org/wiki/%D8%B4%D8%AC%D8%B1%D8%A9_%D9%82%D8%B1%D8%A7%D8%B1.

10. Clustering: An introduction. *A Tutorial on Clustering Algorithms.* http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/

11.جامعة أم القرى. صفات الطالب الناجح. https://old.uqu.edu.sa/aaojaimi/ar/38229

استمارة جمع البيانات

أولاً: البيانات الشخصية :

1) الجنس: ذكر ( ) أنثى ( )

2) العمر: .....

3) القسم : .....

4) العمر عند التسجيل في السنة الأولى : .....

5) تخصص الشهادة الثانوية :

الثانوية الهندسية ( ) ثانوية العلوم الأساسية ( ) الثانوية الإقتصادية ( )

ثانوية علوم الحياة ( ) ثانوية اللغة الإنجليزية ( ) ثانوية اللغة العربية ( )

6) معدل الشهادة الثانوية (بالأرقام) : .....

7) عدد سنوات التأخير في التخرج :

لا يوجد تأخير ( ) سنة ( ) سنتان فما فوق ( )

ثانياً: العوامل التالية تزيد من فرصة نجاح الطالب في المواد الدراسية:

ت	العامل المؤثر	درجة تأثير العامل على نجاح الطالب		
		نعم	نوعاما	لا
1-	الرغبة في المادة			
2-	الصبر			
3-	نسبة الانجاز في تحضير المادة			
4-	التركيز			
5-	الحضور والانضباط			
6-	الاهتمام بالمادة			
7-	القراءة الجماعية			
8-	التنظيم في الوقت			
9-	ايقاف القيد لمدة معينة			
10-	التأخر بالعمر			