الجمهورية العربية السورية
وزارة التعليم العالي
المعهد العالي لإدارة الأعمال

HIGHER INSTITUTE OF BUSINESS ADMINISTRATION

السنة الخامسة

اختصاص إدارة المعطيات وإدارة نظم معلومات

نظام معلومات للمعهد العالي لإدارة الأعمال

(نظام تحليل العلامات)

إعداد الطالب

رامي موقاو

بإشراف الدكتور

د.راتب البلخي

العام الدراسي 2014- 2015

"جميع الآراء الواردة في هذا التقرير تعبر عن وجهة نظر معده ولا يتحمل المعهد أية مسؤولية عن مدى دقة أو مصداقية الآراء المطروحة فيه"

رامي موقاو

الهاتف:0956221486

البريد الإلكتروني:

Rami.Mokao@hotmail.com

الإهداء

إلى القلب الكبير .... إلى رجل بحجم وطن والدي العزيز

إلى من كانت صاحبة القلب الصافي والحنون... إلى من كانت ينبوع الصبر والأمل أمي الحبيبة

إلى أخي العزيز نارت

إلى الذين أحببتهم وأحبوني أصدقائي

إلى الذين أحببتهم وأقدرهم عائلتي

الشكر

أسمى آيات الشكر والاحترام للدكتور راتب البلخي الذي كان عونا لي،

والدكتور طلال عبود الذي أفادنا بغيض من علمه في مادة نظرية القرارات،

والأستاذ محمد عنتور لوقوفه معنا ومساعدته المعطاءة لإتمام هذا العمل،

والأنسة نظرة رحمة من خلال دعمها ومساعدتها لنا،

والكوادر في دائرة الامتحانات لما بذلوه من جهود في إتمام هذا العمل،

وأخيرا نتوجه بالشكر إلى كل المدرسين في المعهد ودائرة شؤون الطلاب ودائرة الامتحانات والمكتبة ولكل من ساهم في إتمام هذا العمل.

الفهرس:

الإهداء 2

الشكر 3

الفهرس: 4

1- ملخص البحث: 7

الفصل التمهيدي: 8

الإطار العام للمشروع 8

1.1- إشكالية البحث: 9

2.1- أهمية البحث: 10

3.1- **أهداف البحث:** 11

4.1- **محددات البحث:** 12

الفصل الثاني: 13

الإطار النظري للمشروع 13

1.2- مقدمة: 14

2.2- مقدمة عن التنقيب: 15

3.2- أنواع تقنيات استخراج البيانات: 17

4.2- تحليل قاعدة البيانات ودعم اتخاذ القرار 18

Database analysis and decision support: 18

5.2- المشاكل التي تقوم بحلها الـ Data Mining: 20

6.2- مهام تقوم بها الـ Data Mining: 22

7.2- خوارزميات الـ Data Mining: 23

1.7.2- Decision Trees: 23

2.7.2- Naïve Bayesian Rule: 23

3.7.2- Clustering: 23

8.2- الأنماط المستخرجة من البيانات: 25

9.2- Classification and Prediction: 26

10.2- القضايا المتعلقة بالتصنيف والتنبؤ: 29

11.2- كيف يتم تقديم نموذج مستمد من البيانات؟ 30

How is the derived model presented?: 30
1.11.2- **قواعد Rules (1Rule)**: 30
2.11.2- **شجرة القرار Decision Trees ID3**: 32
3.11.2- **الصيغ الرياضية Bayesian أو Naïve Bayes's rule**: 37
الفصل الثالث: 40
الإطار العملي للمشروع 40
1.3- تمهيد: 41
2.3- طريقة التحليل والتصميم المستخدمة هي Waterfall: 41
1- **تعريف النطاق (Scope Definition)** : 41
1.1- تعريف حدود ونص المشكلة: 41
2.1- احتياجات قسم إدارة العمليات ونظم المعلومات الإدارية: 42
3.1- المستخدمون: 42
4.1- الهدف: 42
2- **تحليل المشكلة (Problem Analysis)** : 42
1.2- معرفة مجال المشكلة (Understand The Problem Domain): 42
2.2- تحليل المشكلة و الفرص (Analysis Problem and Opportunities): 42
3.2- عمليات الأعمال (Business Process): 42
3- **تحليل المتطلبات (Requirements analysis)**: 48
1.3- المتطلبات الوظيفية (Functional Requirements): 48
2.3- متطلبات غير وظيفية (Non Functional Requirements): 48
4- **تحليل النظم (System Analysis)**: 49
1.4- الفاعلين (Business Actors): 49
2.4- حالات الاستخدام (Use Case): 50
3.4- مخطط حالات الاستخدام (Use Case Diagram): 51
4.4- الوصف النصي لحالات الاستخدام (Business Requirements Use Case Narrative): 52
5.4- مخطط التبعية (Dependency Diagram): 57
6.4- مخطط الصفوف Class Diagram: 58
7.4- مخطط النشاط (Activity Diagram): 59
5- **تصميم النظم (System Design)** : 64

1.5- التكنولوجيا المستخدمة في النظام: (التقرير الفني) : 64
2.5- المخطط الفيزيائي لتدفق البيانات (Physical Data Flow Diagram) PDFD:..... 66
3.5- تصميم نظام تحليل العلامات:..... 66
1- **نظام تحليل علامات الـ PAT** :..... 67
2- **نظام تحليل علامات PAT + 1 2 3 :** 74
3- **نظام تحليل علامات السنوات 1 2 3 + الاختصاص :**..... 87
2- **النتائج:** 92
3- **التوصيات:** 93
4- **المراجع:** 94
5- الملاحق: 95

1- ملخص البحث:

يتناول هذا البحث تحليل وتصميم نظام معلومات للمعهد العالي لإدارة الأعمال،

يقوم هذا النظام:
بتحليل علامات الطلاب الذين تقدموا إلى امتحان القبول، واستنباط قواعد من علامات الطلاب في مواد الاختبار تفيد هذه القواعد في معرفة مستوى قبول الطلاب بالنسبة لعلامة الشهادة الثانوية ومواد اختبار القبول، لوضع معايير لقبول الطلاب استنادا إلى علامة الشهادة الثانوية وعلامات مواد اختبار القبول.

وتحليل علامات الطلاب في مواد اختبار القبول وانعكاسها على مواد السنوات الثلاثة الأولى للطلاب الناجحين في الاختبار والذين سجلوا في المعهد، وتفيد في رفع مستوى اختبار القبول لاختيار طلاب بمستويات أفضل ليتمكنوا من النجاح في مواد السنوات الأولى بشكل أفضل، عن طريق مقارنة مواد اختبار القبول ومواد السنوات الثلاث الأولى.

وتحليل علامات الطلاب في السنوات الثلاثة الأولى وانعكاسها على اختيار الطالب للاختصاص المناسب، وهذا التحليل يفيد في التنبؤ باختصاص الطالب بناء على علاماته ليتم نصح الطالب بالاختصاص المناسب له.

الفصل التمهيدي:

الإطار العام للمشروع

1.1- إشكالية البحث:

في ظل كثرة عدد الطلاب ووجود الكثير من التطورات والتعديلات مما يصعب عملية تحليل بيانات الطلاب يدويا.

تتلخص مشكلة البحث بالأسئلة التالية:

1- هل توجد معايير واضحة عند الطالب لاختيار الاختصاص المناسب له؟

2- هل يوجد نظام معلومات لتحليل علامات الطلاب، ولاتخاذ إجراءات (قرارات) بعد التحليل؟

3- هل يمكن أرشفة نتائج التحليل اليدوي، والاطلاع عليها من قبل أية جهة؟

2.1- أهمية البحث:

تتجلى أهمية البحث من خلال:

1- تحليل علامات الطلاب الذين تقدموا إلى اختبار القبول، واستنباط قواعد من علامات الطلاب في مواد الاختبار تفيد هذه القواعد في معرفة مستوى قبول الطلاب بالنسبة لعلامة الشهادة الثانوية.

2- تحليل علامات الطلاب في مواد اختبار القبول وانعكاسها على مواد السنوات الثلاثة الأولى للطلاب الناجحين في الاختبار والذين سجلوا في المعهد، تفيد في رفع مستوى اختبار القبول لاختيار طلاب بمستويات أفضل ليتمكنوا من النجاح في مواد السنوات الثلاثة الأولى بشكل أفضل، عن طريق مقارنة مواد اختبار القبول ومواد السنوات الثلاث الأولى.

3- تحليل علامات الطلاب في السنوات الثلاثة الأولى وانعكاسها على اختيار الطالب للاختصاص المناسب، وهذا التحليل يفيد بالتنبؤ باختصاص الطالب بناء على علاماته ونصح الطالب بالاختصاص المناسب له.

وبالتالي يسهم هذا البحث في حل مشكلة اختيار الطالب للاختصاص المناسب له، ويسهل معرفة الطالب إذا كان قادرا على النجاح في المرحلة التحضيرية الأولى.

3.1- أهداف البحث:

تتلخص أهداف البحث بما يلي:

1- تحليل علامات الطلاب المتقدمين لاختبار القبول.

2- وضع معايير لقبول الطلاب استنادا إلى علامة الشهادة الثانوية.

3- تحليل علامات الطلاب في السنوات الثلاثة الأولى.

4- رفع مستوى اختبار القبول لاختيار طلاب بمستويات أفضل ليتمكنوا من النجاح في مواد السنوات الثلاثة الأولى بشكل أفضل، بناء على مقارنة علامات الطلاب في مواد اختبار القبول ومواد السنوات الثلاثة الأولى.

5- تقديم النصح للطالب بالاختصاص المناسب له بناء على علاماته في السنوات الثلاثة الأولى.

6- أرشفة النتائج بوضع قواعد واضحة (Rules) يتم وضع المعايير بناء عليها.

4.1- محددات البحث:

1- الطلاب الذين تقدموا لاختبارات القبول (الناجحين – احتياط – مرفوضين) من العام 2009 إلى العام 2014.
2- طلاب السنوات الثلاثة الأولى من العام 2009-2010 إلى العام 2013-2014.
3- طلاب الاختصاصات (إدارة نظم المعلومات – الموارد البشرية – المالية – التسويق) للسنوات الرابعة والخامسة من العام 2012-2013 إلى العام 2014-2015.

الفصل الثاني:

الإطار النظري للمشروع

1.2- مقدمة:

التنقيب في كميات كبيرة من البيانات لتحويلها إلى معلومات ومعارف مفيدة للوصول إلى القرار أو استخدامها في تحليل السوق ومراقبة الإنتاج والاستكشافات العلمية.

واستخدام خوارزميات الـ Data Mining (استخراج البيانات) لاحتواء المشروع على تطبيقات التصنيف والتنبؤ، سنعتمد على الـ (Classification and Prediction) وهي من مهام الـ Data Mining التي تتم بخطوتين (بناء النموذج، واستخدام النموذج).

وكذلك سنعرض أمثلة محلولة يدويا عن الخوارزميات (لعدد صغير من العينات) لإيضاح كيفية عمل كل خوارزمية على حدى.

2.2- مقدمة عن التنقيب:[1]

- ما هو التنقيب (Mining)؟

عمل من أعمال الحفر في الأرض حيث يمكن استخراج الخامات والمعادن.

- ما هي التنقيب في البيانات (Data Mining)؟

هي استخراج (التنقيب) المعرفة من كميات كبيرة من البيانات.

- ماذا يحفز التنقيب في البيانات؟ لماذا هو مهم؟

لقد اجتذب استخراج البيانات قدرا كبيرا من الاهتمام في مجال صناعة المعلومات والمجتمع ككل في السنوات الأخيرة، ويرجع ذلك إلى توافر كميات هائلة من البيانات والحاجة لتحويل هذه البيانات إلى معلومات ومعارف مفيدة.

عملية التنقيب عن المعلومات (Data Mining) لاكتشاف المعارف: هي عبارة عن تطبيق خوارزميات معينة لاستخلاص نماذج معطيات مفيدة في ايجاد أنماط (Patterns) موجودة في البيانات، وبالتالي فإن عملية التنقيب عن المعطيات هي اكتشاف المعلومات المخبأة، وعادة تكون هذه المعلومات غير معروفة وتتألف من كمية كبيرة من المعطيات، ويكون الهدف هو اكتشاف العلاقات الهامة بين المعطيات لإيجاد المعلومة المطلوبة، كما أن عملية التنقيب عن المعطيات تلعب دورا هاما في عملية دعم القرار أو الوصول إلى قرار.

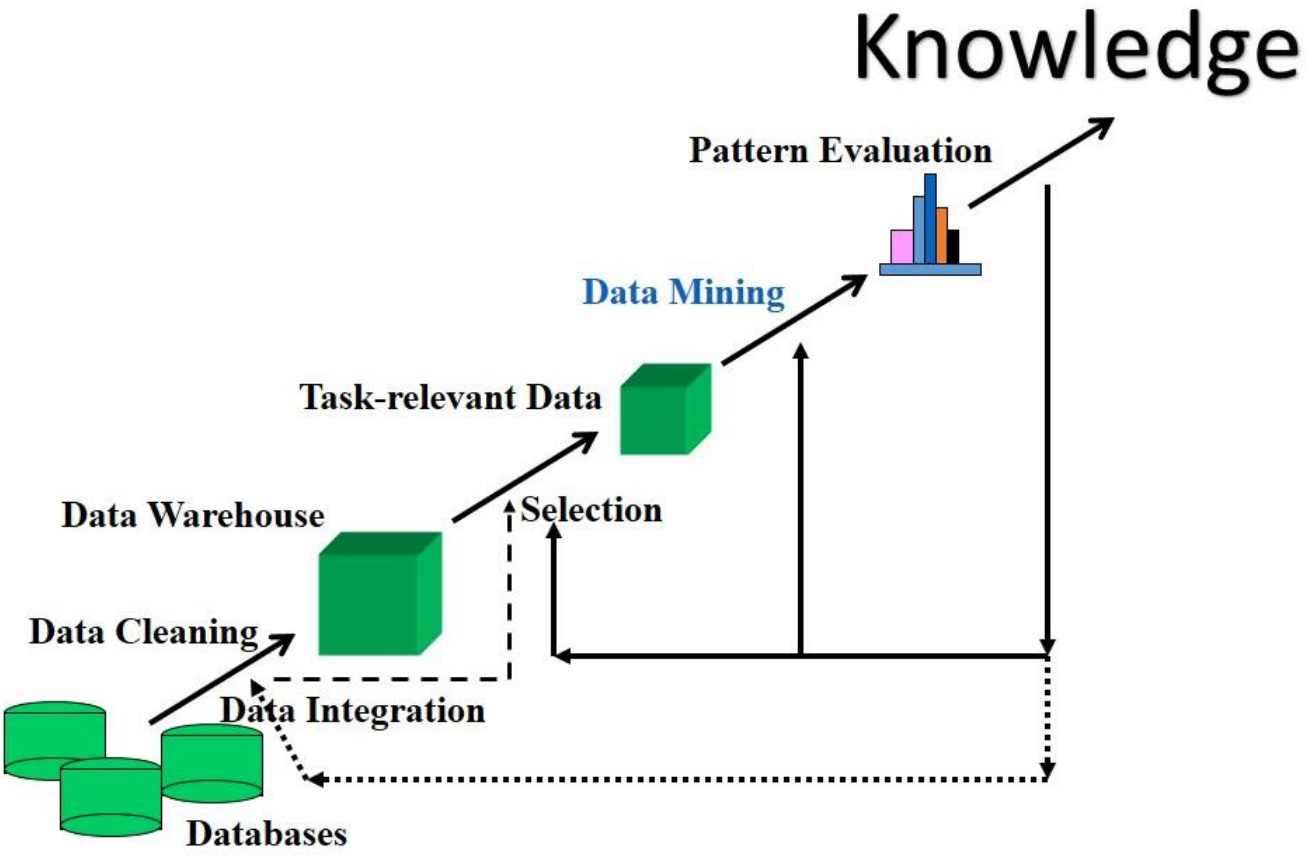
المعلومات والمعارف المكتسبة يمكن استخدامها لتطبيقات تستخدم في تحليل السوق، والكشف عن الغش، والمحافظة على العملاء، لمراقبة الإنتاج والاستكشافات العلمية، والتنبؤ.

استخراج البيانات (Data Mining) هو مفيد في إيجاد أنماط خفية من مجموعة بيانات من خلال تحليل الارتباطات بين القيم.

[1] Data Mining Concepts and Techniques, Jiawei Han and Micheline Kamber, 2006

- التنقيب عن البيانات هي خطوة ضمن عملية اكتشاف المعرفة -

Data mining as a step in the process of knowledge discovery:



3.2- أنواع تقنيات استخراج البيانات:[2]

أ- التنقيب الموجه (Supervised):

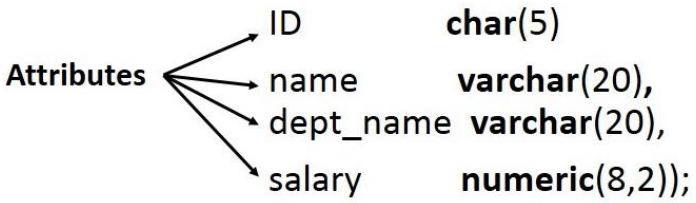
استخراج البيانات يتطلب من المستخدم لتحديد واصفة الهدف (attribute) ومجموعة من واصفات(attributes) الإدخال،

وتشمل خوارزميات الـ Supervise للتنقيب عن البيانات:

- قواعد Rules (1R)
- شجرة القرار Decision Trees ID3
- الصيغ الرياضية Bayesian و Naïve Bayes's rule
- الشبكات العصبية Neural Networks

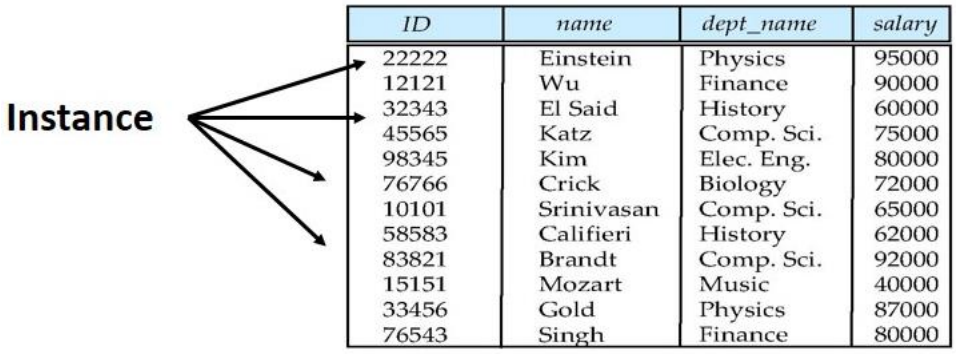
السمة أو الواصفة (attributes):

هي نوعية البيانات التي يمكن اعدادها للتسجيل والتحليل.



وكل واصفة تحوي عدد من الـ instance:

Instance: هي القيم أو المتغيرات التي توجد في الواصفات.



ID	name	dept_name	salary
22222	Einstein	Physics	95000
12121	Wu	Finance	90000
32343	El Said	History	60000
45565	Katz	Comp. Sci.	75000
98345	Kim	Elec. Eng.	80000
76766	Crick	Biology	72000
10101	Srinivasan	Comp. Sci.	65000
58583	Califieri	History	62000
83821	Brandt	Comp. Sci.	92000
15151	Mozart	Music	40000
33456	Gold	Physics	87000
76543	Singh	Finance	80000

[2] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

ب- التنقيب غير الموجه (Unsupervised):

تقنية استخراج البيانات لا يجب أن تتنبأ بالواصفة (attributes)، التجميع (Clustering) هو مثال جيد لـ Unsupervised Data Mining وهي وضع مجموعة نقاط البيانات غير المتجانسة إلى مجموعات فرعية بحيث تكون نقاط البيانات في كل مجموعة فرعية هي أكثر أو أقل تجانسا.

4.2- تحليل قاعدة البيانات ودعم اتخاذ القرار [3]

Database analysis and decision support:

1- تحليل وإدارة السوق - Market analysis and management:

- مصدر المعلومات لعملية التحليل من:

عمليات بطاقات الائتمان Credit card transactions، بطاقات الولاء loyalty cards، كوبونات الحسم discount coupons، مكالمات شكاوي الزبائن customer complaint calls، أسلوب حياة العامة plus (public) lifestyle studies.

- الهدف التسويقي:

إيجاد (Clusters) مجموعات من نماذج الزبائن الذين يتشاركون نفس الشخصية (الاهتمام – مستوى الدخل – عادات الانفاق – .....).

- تنميط العملاء - Customer profiling :

استخراج البيانات يمكن ان تحدد ما هي أنواع زبائن الشراء وما هي المنتجات (Clustering – Classification).

- تحليل السلة التسويقية – Market basket analysis.
- تقاطع المبيعات - Cross selling.
- تقسيم السوق – Market segmentation.

[3] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

2- تحليل وإدارة المخاطر - Risk analysis and management:

- وضع خطط مالية وتقييم الأصول.
- تحليل وتنبؤ بالتدفقات المالية.
- التنبؤ - Forecasting.
- المحافظة على العملاء - customer retention.
- التحكم بالجودة - quality control.
- تحليل التنافس - competitive analysis.

3- كشف وإدارة الاحتيال - Fraud Detection and Management:

- التطبيق:

يستخدم بشكل واسع في العناية الصحية، محلات بيع التجزئة، بطاقات الائتمان، الاتصالات، ....

- أسلوب العمل:

استخدام البيانات التاريخية لبناء نماذج للسلوك الاحتيالي واستخدام استخراج البيانات للمساعدة في تحديد حالات مماثلة.

- أمثلة:

التلاعب بالتأمين، غسيل الأموال، التلاعب بالتأمين الصحي، ...

5.2- المشاكل التي تقوم بحلها الـ Data Mining:[4]

1.5.2- ما هي المنتجات التي تشترى مع الكاميرا الرقمية؟

بناء على المشتريات السابقة (Shopping cart).

على سبيل المثال، إذا اشتريت كاميرا رقمية، يتم أيضا شراء كرت ذاكرة، والبطارية، والطابعة.

وهناك أسطورة معروفة وفي وول مارت(Walmart)، الرجال في الـ 35 من عمرهم يشترون البيرة والحفاضات معا في يوم الجمعة.

--> هنا نستخدم الـ Association Analysis

أسئلة مشابهة:

ما هي المنتجات التي ينصح بها في المتاجر الالكترونية (Online stores) كـ Amazon.com، متاجر الأفلام، .... ؟

ما هي المواد التي يجب أن تعرض معا عند التاجر؟

2.5.2- هل الطالب سينجح ويدخل الجامعة؟

بناء على الجنس، دخل الأب، مستوى الذكاء، علامات الثانوية .....

على سبيل المثال:

[مستوى دخل الأب > 35000 و مستوى الذكاء >100] ← [الجامعة = ناجح]

--> هنا نستخدم الـ Classification (Prediction)

أسئلة مشابهة:

هل هذا البريد الإلكتروني غير مرغوب؟ (spam filtering).

كيف هي الخدمة الائتمانية الخاصة بك جيدة / سيئة ؟ (credit scoring).

تمييز الرسائل المكتوبة بخط اليد؟ (pen recognition).

ما هو هذا الجين (علم أحياء)؟ (bioinformatics)

هل هذا الشخص يتصرف وكأنه إرهابي؟

---

[4] Data Mining With Microsoft SQL Server 2008, Jamie MacLennan – ZhaoHui Tang – Bogdan Crivat, 2009

3.5.2- ما هو عمر الشخص؟

بناء على الهوايات، الحالة العائلية، عدد الأولاد، الدخل، عدد السيارات، ....

--> هنا نستخدم الـ Regression (Prediction)

أسئلة مشابهة:

ما هي كمية المبيعات من الآيس كريم في الشهر القادم؟ (sales prediction)

ما هو سعر الأسهم في الأسبوع المقبل؟ (stock prediction)

ما هو دخل العملاء؟ (marketing)

ما هو زمن حدوث خطأ في البرمجية (خطأين في السنة) ؟ (bug tracking)

4.5.2- من هم زوار موقع الويب الالكتروني؟

بناء على تحديد مجموعات مماثلة على أساس التركيبة السكانية، وأنماط الزيارة.

على سبيل المثال: قراء الأخبار اليومية، مستخدمي البريد الإلكتروني، والمتسوقين،...

--> هنا نستخدم الـ Segmentation (clustering)

أسئلة مشابهة:

تحديد مجموعات من الجينات (علم أحياء)؟ (bioinformatics)

تحديد مجموعات من مواقع إصابات الكوليرا في لندن؟ (spatial data mining)

تحديد مجموعة الزبائن للمتاجر كـ (Amazon،E-Bay، MSN، WalMart، ....) ؟ (target marketing)

تحديد مجموعات من الوثائق؟ (text categorization)

5.5.2- يمكن أن تكون هذه الحزمة الشبكية هجوم من فيروس؟

بناء على توقع احتمال نمط حزمة شبكة الاتصال.

--> هنا نستخدم الـ Anomaly detection (outlier detection)

أسئلة مشابهة:

هل نتائج تحليل المشفى عادية؟ (Adverse drug effect detection)

هل هذه العمليات الائتمانية (credit transaction) مزورة؟ (fraud detection)

هل هذا الشخص يتصرف بشكل غير عادي، وربما يستحق مستوى عال من الاهتمام الأمني؟

6.2- مهام تقوم بها الـ Data Mining:[5]

Classification

Regression

Segmentation (Clustering)

Association Analysis

Anomaly detection

Sequence Analysis

Time-series Analysis

Text categorization

Advanced insights discovery

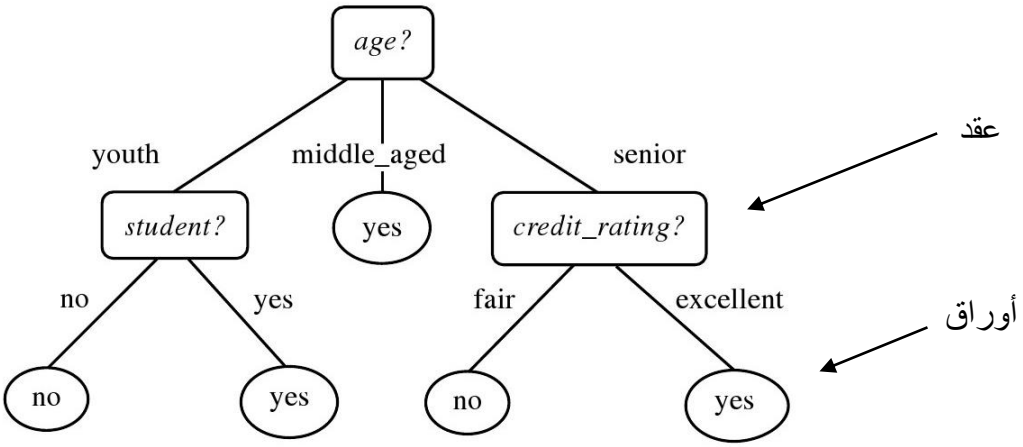
...Others

[5] Data Mining With Microsoft SQL Server 2008, Jamie MacLennan – ZhaoHui Tang – Bogdan Crivat, 2009

7.2- خوارزميات الـ Data Mining:[6]

1.7.2- Decision Trees:

شجرة القرار هي بنية شجرية تشبه المخطط، حيث كل عقدة داخلية (عقدة غير طرفية) ترمز إلى اختبار الواصفة (attribute)، يمثل كل فرع نتائج الاختبار، وكل عقدة ورقة (أو عقدة طرفية) تحمل تسمية الطبقة، العقدة الأكثر علوا في الشجرة هي العقدة الجذر.



2.7.2- Naïve Bayesian Rule:

ويمكن أن تتوقع احتمال العضوية (ورود – تكرار) في الـ Transaction، مثل احتمال أن ينتمي غرض معين إلى عملية الشراء.

ويستند تصنيف النظرية الافتراضية على نظرية بايز (Bayes) الاحتمالية.

3.7.2- Clustering:

هذه العملية لتجميع مجموعة من الأشياء المادية أو المجردة إلى طبقات من أشياء متماثلة.

وهي تقوم بتجميع الكائنات من البيانات التي تشبه بعضها البعض ضمن نفس المجموعة، وهي تختلف عن الكائنات في المجموعات الأخرى.

---

[6] Data Mining With Microsoft SQL Server 2008, Jamie MacLennan – ZhaoHui Tang – Bogdan Crivat, 2009

وهناك مجموعة من كائنات البيانات يمكن علاجها بشكل جماعي كمجموعة واحدة وهكذا يمكن اعتباره شكلا من أشكال ضغط البيانات أو توزيع البيانات.

على الرغم من أن التصنيف (clustering) هو وسيلة فعالة للتمييز بين مجموعات أو فئات، فإنه يتطلب أحيانا جمع عدد كبير من المجموعات والأنماط، التي تستخدم لنمذجة كل مجموعة.

غالبا ما يكون مرغوبا أكثر للمشروع أن يتم في الاتجاه المعاكس: أولا تقسيم مجموعة من البيانات إلى مجموعات على أساس تشابه البيانات (using clustering)، ومن ثم وضع تسميات لعدد صغير نسبيا من المجموعات.

مزايا إضافية لهذه العملية أنها قابلة للتكيف مع التغيرات وتساعد على تمييز المجموعات المختلفة.

4.7.2- Sequence Clustering

5.7.2- Association Rules

6.7.2- Neural Network

7.7.2- Time Series

8.7.2- Support Vector Machines ...

8.2- الأنماط المستخرجة من البيانات:[7]

- تنقيب (استخراج) البيانات (Data Mining) - ما هي أنواع الأنماط (Patterns) التي يمكن التنقيب عنها؟

تستخدم وظائف استخراج البيانات لتحديد أنواع خاصة من الأنماط التي يمكن العثور عليها في مهام استخراج البيانات.

بشكل عام، مهام استخراج البيانات التي يمكن تصنيفها إلى فئتين:

وصفية (Descriptive) - تنبؤية (Predictive) .

مهام التنقيب الوصفية (Predictive) تميز الخصائص العامة للبيانات في قاعدة البيانات.

مهام التنقيب التنبؤية (Descriptive) تقوم بالاستدلال على البيانات الحالية من أجل الحصول على التوقعات.

- هل كل الأنماط (Patterns) مفيدة ومثيرة للاهتمام؟

نظام استخراج البيانات لديه القدرة على توليد الآلاف أو حتى الملايين من الأنماط، أو القواعد.

فقط جزء صغير من الأنماط التي ولدت ستكون في الواقع مفيدة للمستخدم.

وهذا يثير بعض التساؤلات لاستخراج البيانات:

ما الذي يجعل النمط مثيرا للاهتمام أو مفيدا للمستخدم؟

يمكن لنظام استخراج البيانات توليد كل الأنماط؟

يمكن لنظام استخراج البيانات توليد أنماط مفيدة فقط؟

للإجابة عن **السؤال الأول**، النمط يكون مفيد للمستخدم إذا كان:

1- سهل الفهم من قبل البشر.

2- ساري المفعول على البيانات الجديدة مع درجة من اليقين certainty.

3- يحتمل أن يكون مفيدا.

4- يمكن روايتها (بشكل نصي مفهوم).

النمط (Patterns) المفيد الذي يؤكد نظرية معينة ويصبح قاعدة معرفية.

[7] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

والسؤال الثاني يشير إلى فعالية خوارزمية استخراج البيانات:
غالبا توليد كل الأنماط المحتملة من قبل الـ Data Mining تكون غير واقعية وغير فعالة.
بدلا من ذلك القيود المتوفرة للمستخدم والتدابير ينبغي أن تستخدم لتركيز البحث عن الأنماط.

والسؤال الثالث هي مشكلة التحسين في استخراج البيانات:
ومن المرغوب فيه جدا لأنظمة استخراج البيانات توليد أنماط مفيدة فقط.
وهذا من شأنه أن يكون أكثر كفاءة للمستخدمين ولأنظمة استخراج البيانات.
إن تحديد عدد الأنماط المفيدة تعتبر ضرورية لاكتشاف كفاءة الأنماط المفيدة للمستخدم.
هذا العدد من الأنماط يمكن أن يستخدم بعد خطوة استخراج البيانات من أجل ترتيب الأنماط المكتشفة وتصفية (فلترة) الأنماط غير المفيدة.
الأهم من ذلك، يمكن استخدام عدد الأنماط لتوجيه وتقييد عملية الاكتشاف، وتحسين الكفاءة في البحث عن طريق اختصار الأنماط الفرعية المكررة (المحددة مسبقا).

9.2- Classification and Prediction:[8]

سنتوسع بالشرح عن الـ Classification and Prediction لاحتواء المشروع على تطبيقات التصنيف والتنبؤ:
التصنيف هو عملية إيجاد نموذج (أو وظيفة) يصف ويميز فئات البيانات أو المفاهيم، لغرض التمكن من استخدام نموذج التنبؤ لفئة من البيانات غير معروفة التسمية والطبقة.
ويقوم النموذج المستمد على تحليل مجموعة من البيانات.

[8] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

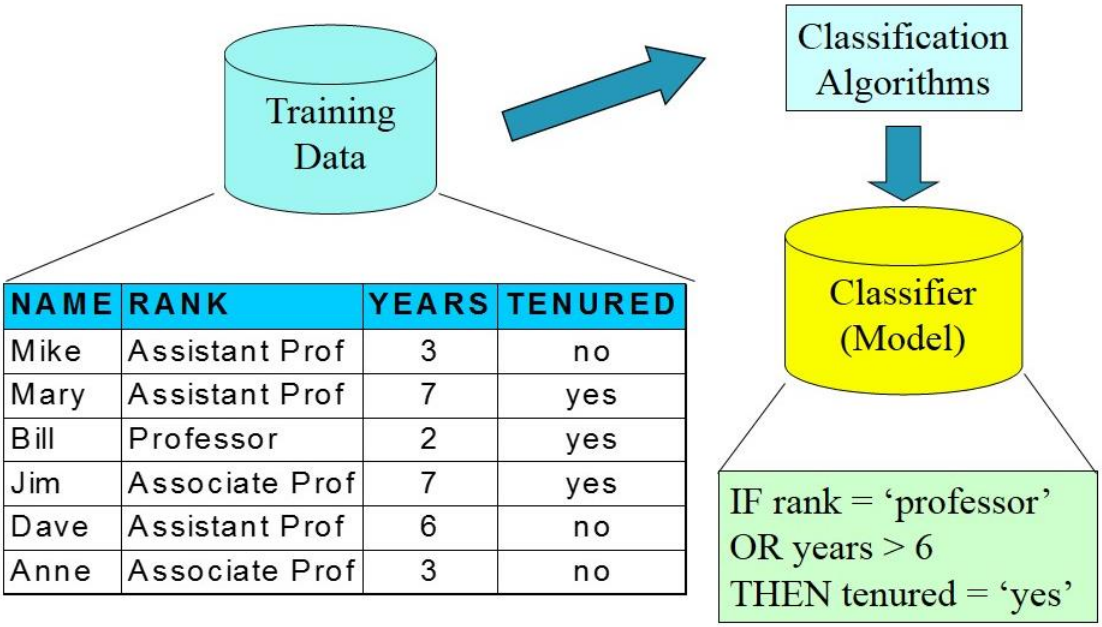
عملية التصنيف تتم بخطوتين:

أ- **بناء نموذج** : وصف مجموعة من الفئات المحددة سلفا.

ويفترض أن كل عينة تنتمي إلى فئة محددة مسبقا، على النحو الذي يحدده تسمية طبقة الواصفة (attribute).

ويمثل هذا النموذج: قواعد التصنيف (1Rule)، وشجرة القرار (Decision Tree ID3)،

أو الصيغ الرياضية (Naïve Bayes's rule).



NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

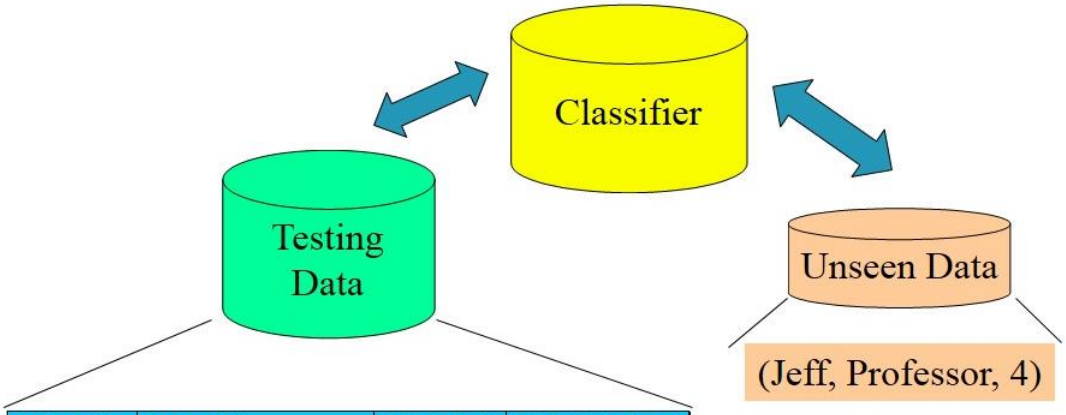
الخطوات:

أولا: تطبيق الخوارزمية على البيانات التجريبية.

ثانيا: استخلاص نماذج التصنيف.

ب- **استخدام النموذج:** لتصنيف المستقبل والأشياء غير المعروفة.

- تقدير دقة النموذج.
- مقارنة عينة الاختبار مع نتائج النموذج.
- معدل الدقة هو النسبة المئوية لمجموعة عينات الاختبار التي تصنف بشكل صحيح عن طريق النموذج.
- اختبار مجموعة مستقلة من العينة، وإلا سيحدث إفراط في عدد الأصناف.



NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes

الخطوات:

أولا: مقارنة البيانات الجديدة مع نموج التصنيف.

ثانيا: اعتماد النتيجة.

ثالثا: إضافة النتيجة للنموذج.

10.2- القضايا المتعلقة بالتصنيف والتنبؤ:[9]

1- إعداد البيانات:

- تنظيف البيانات (Data cleaning):

تحضير البيانات لتخفيف تأثير وتبعات القيم الخاطئة أو الخالية.

- تحليل الصلة ((Relevance analysis (feature selection):

إزالة الواصفات (attributes) التي ليس لها صلة والزائدة عن الحاجة.

- تحويل البيانات (Data transformation):

تحويل البيانات إلى أشكال معروفة ومرتبة.

2- تقييم طرق التصنيف:

- دقة التنبؤ (Predictive accuracy).
- سرعة وقابلية (Speed and scalability):

الوقت لبناء نموذج.

الوقت لاستخدام نموذج.

- قوة البنية (Robustness):

التعامل مع القيم الخاطئة والمفقودة.

- التدرجية (Scalability):

الكفاءة في قواعد البيانات الموجودة على القرص.

- قابلية التفسير (Interpretability):

الفهم والاستيعاب المقدم من قبل النموذج.

- الفوائد من القواعد (Goodness of rules):

حجم شجرة القرار.

ترتيب قواعد التصنيف.

[9] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

11.2- كيف يتم تقديم نموذج مستمد من البيانات؟

How is the derived model presented?:[10]

النموذج المستمد يكون مشكلا من خوارزميات مختلفة، وعادة تكون النتائج متقاربة جدا أو تكون نفس النتائج.

مثل:

1.11.2- قواعد Rules (1Rule):

هناك أنواع كثيرة من البنية (simple structure):

- واصفة (attribute) واحدة تقوم بكل العمل (الاعتماد على واصفة واحدة).
- كل الواصفات (attributes) تساهم وبالتساوي وبشكل مستقل في العملية.
- المزيج الخطي (linear combination) الموزون قد يقوم بالعملية.
- الأساس قائم على استخدام عدد قليل من النماذج.

سنحل مثالا بشكل يدوي لكمية قليلة من البيانات، لأننا لا نستطيع حل مثال معقد لطول وصعوبة الحل:

مثال: (مثال مشهور في الـ Data Mining)

بيانات ملعب تنيس تمثل الطقس والحرارة وإمكانية اللعب:

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

[10] Data Mining Concepts and Techniques, , Jiawei Han and Micheline Kamber, 2006

Attribute	Rules	Errors	Total errors
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temp	Hot → No	2/4	5/14
	Mild → Yes	2/6	
	Cool → Yes	1/4	
Humidity	High → No	3/7	4/14
	Normal → Yes	1/7	
Windy	False → Yes	2/8	5/14
	True → No	3/6	

نختار إحدى القواعد التي فيها أقل الأخطاء:

Attribute	Rules	Errors
Outlook	Sunny → No	2/5
	Overcast → Yes	0/4
	Rainy → Yes	2/5

إذا ورد لدينا يوم جديد (غير موجود في الجدول السابق):

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

حسب القاعدة المختارة:

Outlook = Sunny → No

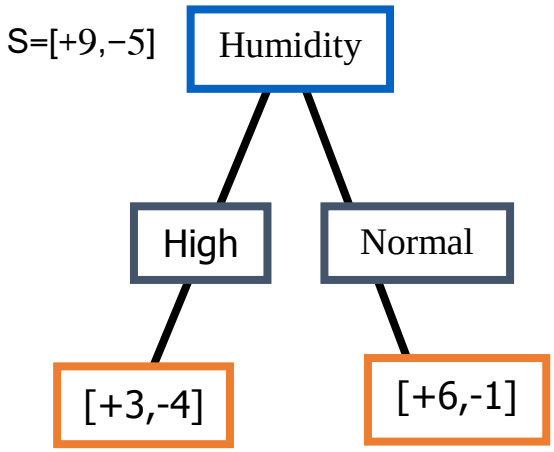
2.11.2- شجرة القرار Decision Trees ID3:

مثال: التنس السابق:

1- نحسب الـ no والـ yes لكل عقدة وتفرعها:

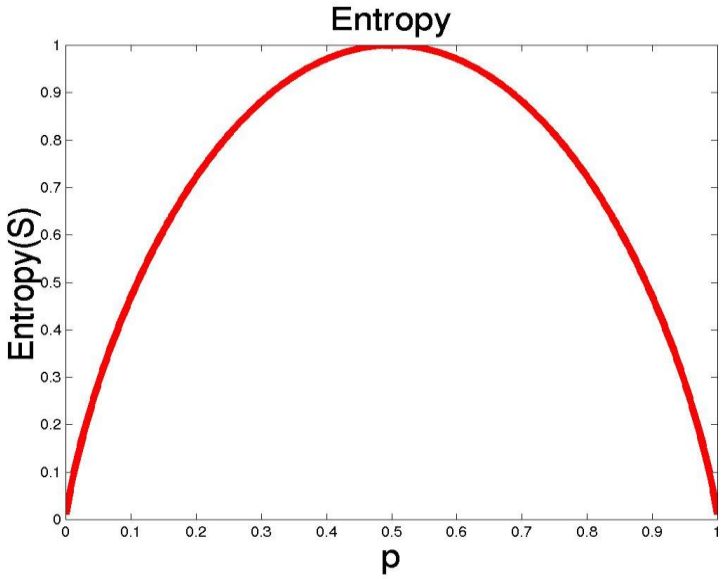
حيث أن الـ Yes (+)، والـ No (−).

- عقدة الـ Humidity:



نحسب الـ Entropy لكل ورقة:

$$Entropy(S) = - p+ \log_2 p+ - p- \log_2 p-$$



S هي عينة من الأمثلة التدريبية.

+P هي عدد الأمثلة الإيجابية.

−P هو عدد الأمثلة السلبية.

Entropy يقيس اختلاط +P و −P.

Entropy (High) = 0.985

Entropy (Normal) = 0.592

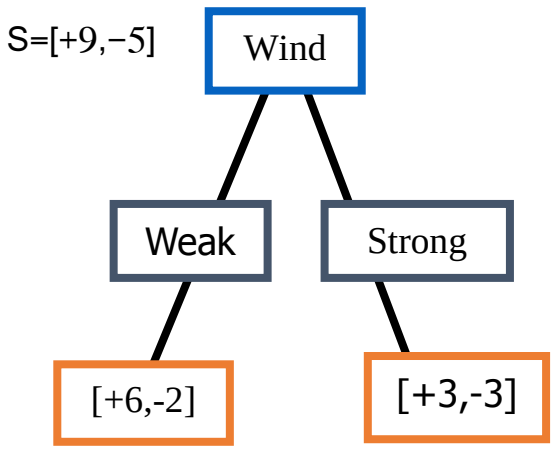
Entropy (Humidity) = 0.940

نحسب الـ Gain لكل ورقة:

$$Gain(S,A) = Entropy(S) - \sum_{v \in values(A)} |S_v|/|S| \; Entropy(S_v)$$

$$Gain(S, Humidity) = 0.940 - (7/14)*0.985 - (7/14)*0.592$$
$$= 0.151$$

- عقدة الـ Wind:



نحسب الـ Entropy لكل ورقة:

Entropy (Weak) = 0.811

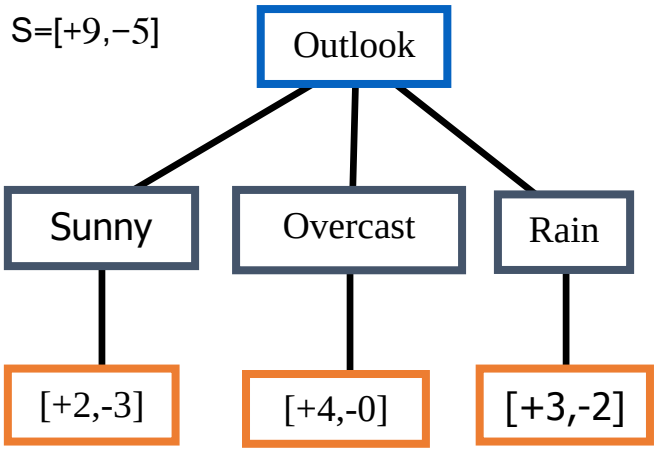
Entropy (Strong) = 1

Entropy (Windy) = 0.940

نحسب الـ Gain لكل ورقة:

$$Gain(S, Wind) = 0.940 - (8/14)*0.811 - (6/14)*1$$
$$= 0.048$$

- عقدة الـ Outlook:



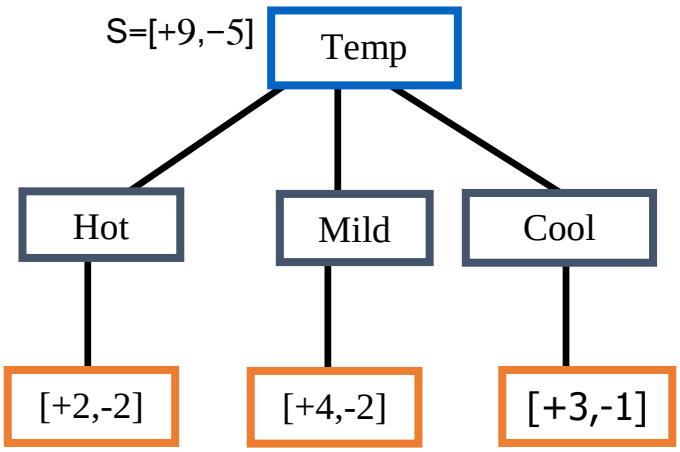
نحسب الـ Entropy لكل ورقة:

$$Entropy(Sunny) = 0.971$$
$$Entropy(Overcast) = 0$$
$$Entropy(Rain) = 0.971$$
$$Entropy(Outlook) = 0.940$$

نحسب الـ Gain لكل ورقة:

$$Gain(S, Outlook) = 0.940 - (5/14)*0.971 - (4/14)*0 - (5/14)*0.971$$
$$= 0.247$$

- عقدة الـ Temp:



نحسب الـ Entropy لكل ورقة:

Entropy (Hot) = 1

Entropy (Mild) = 0.918

Entropy (Cool) = 0.811

Entropy (Temp) = 0.940

نحسب الـ Gain لكل ورقة:

$$Gain(S , Temp) = 0.940 - (4/14)*1 - (6/14)*0.918 - (4/14)*0.811$$
$$= 0.029$$

نرتب الـ Gain تنازليا:

Gain(S , Outlook) = 0.247

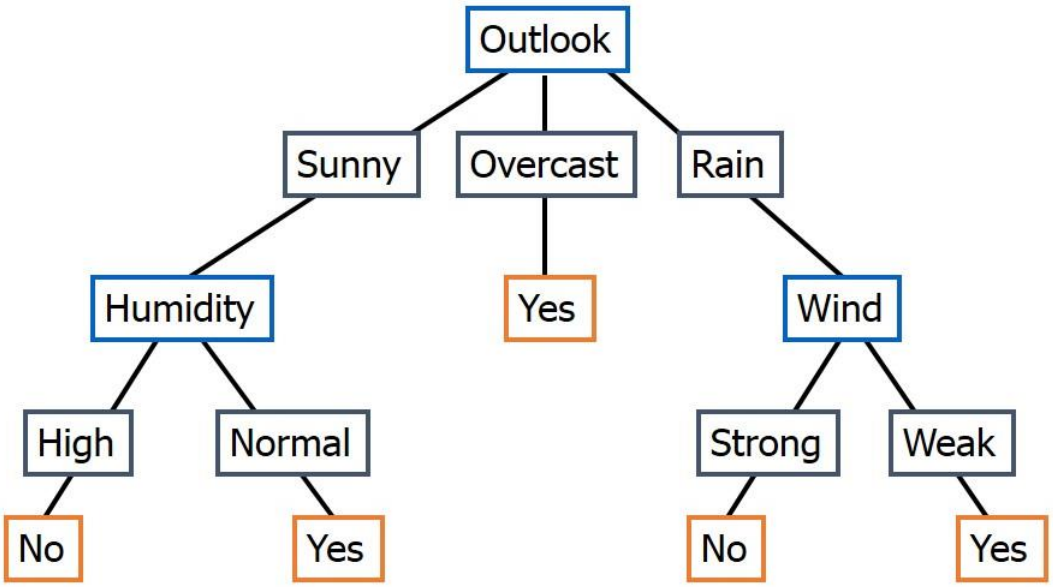
Gain(S , Humidity) = 0.150

Gain(S , Wind) = 0.048

Gain(S , Temp) = 0.029

نرتب الشجرة حسب الـ Gain الأعلى:

(مع إهمال الـ Temp لأنه الأصغر من حيث الـ Gain)



من خلال الشجرة نصيغ القواعد التالية:

R_1: If (Outlook=Sunny) $\wedge$ (Humidity=High) Then PlayTennis=No

R_2: If (Outlook=Sunny) $\wedge$ (Humidity=Normal) Then PlayTennis=Yes

R_3: If (Outlook=Overcast) Then PlayTennis=Yes

R_4: If (Outlook=Rain) $\wedge$ (Wind=Strong) Then PlayTennis=No

R_5: If (Outlook=Rain) $\wedge$ (Wind=Weak) Then PlayTennis=Yes

عند ورود بيانات جديدة:

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

هنا نختار الـ NO

3.11.2- الصيغ الرياضية Bayesian أو Naïve Bayes's rule:

عكس خوارزمية 1R تستخدم كل الواصفات (attributes).

يوجد افتراضين: الواصفات:

متساوية الأهمية.

مستقلة إحصائيا (نظرا لقيمة الصف).

افتراض الاستقلالية بين الواصفات تعطي أحيانا قيم غير صحيحة.

مثال: التنس السابق:

Outlook			Temperature			Humidity			Windy			Play	
	Yes	*No*		*Yes*	*No*		*Yes*	*No*		*Yes*	*No*	*Yes*	*No*
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

عند ورود بيانات جديدة:

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

نحسب الاحتمال H (yes/no) باستخدام قانون Bayes الاحتمالي:

احتمال حدوث الحدث H علما بأن الدليل E قد وقع:

$$\Pr[H \mid E] = \frac{\Pr[E \mid H]\Pr[H]}{\Pr[E]}$$

الدليل E هو instance (Evidence E = instance).

الحدث H هو قيمة فئة (class) من الـ instance (Event H = class value for instance).

قسم Naïve Bayes الدليل E في القانون لأجزاء (أدلة E1، E2، E3، ...):

$$\Pr[H \mid E] = \frac{\Pr[E_1 \mid H]\Pr[E_2 \mid H]\ldots\Pr[E_n \mid H]\Pr[H]}{\Pr[E]}$$

إذا كانت الحرارة بالدرجات نحسب بالقانون:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

حساب احتمال الـ yes:

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

← ***Evidence E***

$$\Pr[yes \mid E] = \Pr[Outlook = Sunny \mid yes]$$
$$\times \Pr[Temperature = Cool \mid yes]$$
$$\times \Pr[Humidity = High \mid yes]$$
$$\times \Pr[Windy = True \mid yes]$$
$$\times \frac{\Pr[yes]}{\Pr[E]}$$

Probability of class "yes"

$$= \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{\Pr[E]}$$

حساب احتمال الـ no:

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

← ***Evidence E***

$$\Pr[no \mid E] = \Pr[Outlook = Sunny \mid no] \times \Pr[Temperature = Cool \mid no] \times \Pr[Humidity = High \mid no] \times \Pr[Windy = True \mid no] \times \frac{\Pr[no]}{\Pr[E]}$$

Probability of class "no"

$$= \frac{\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}}{\Pr[E]}$$

$$\Pr[yes \mid E] = 0.0053$$

$$\Pr[no \mid E] = 0.0206$$

P(yes) = 0.0053 / (0.0053 + 0.0206) = 0.205

P(no) = 0.0206 / (0.0053 + 0.0206) = 0.795

هنا نختار الـ NO

4.11.2- الشبكات العصبية Neural Networks.

نجد أن الجواب بالخوارزميات الأربعة عن بيانات جديدة غير موجودة: هي No

Outlook	Temp	Humidity	Windy	Play
Sunny	Cool	High	True	?

الفصل الثالث:

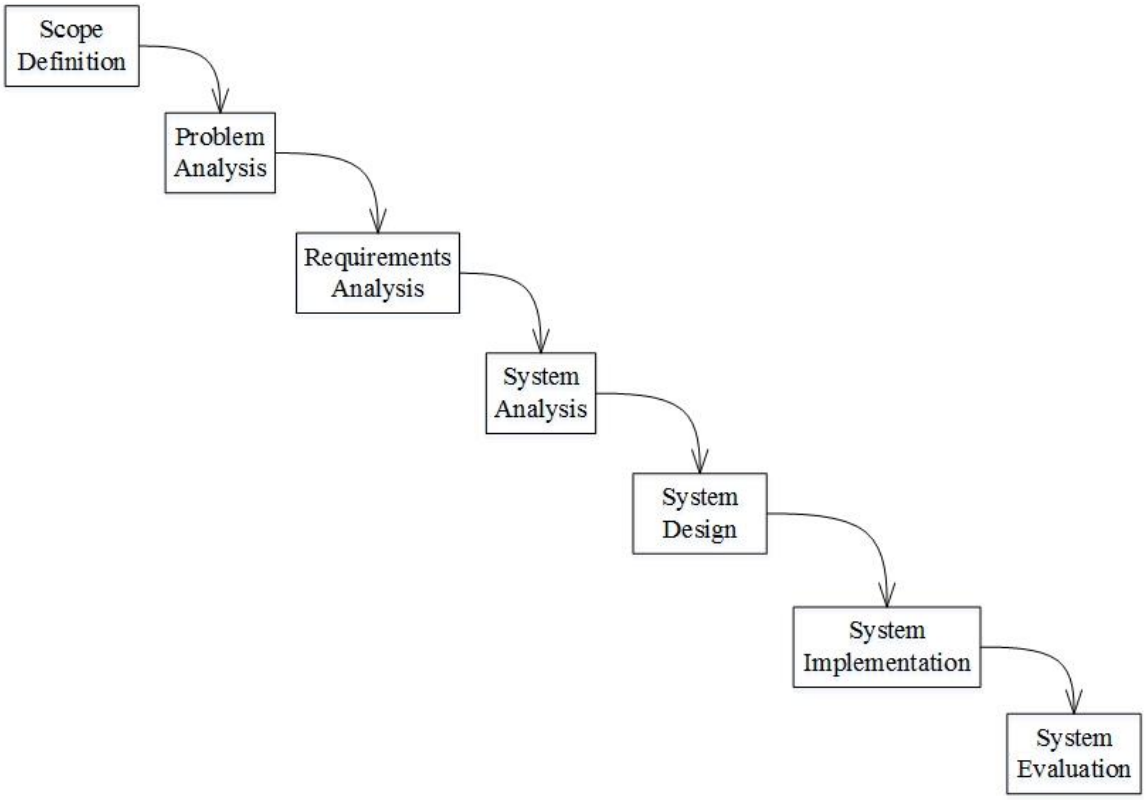
الإطار العملي للمشروع

1.3- تمهيد:

سنقوم ببناء نظام معلومات للمعهد العالي لإدارة الأعمال، نظام تحليل علامات لقسم إدارة العمليات ونظم المعلومات الإدارية.
حيث تم اختيار طريقة بناء النظام وهي طريقة (Waterfall).

2.3- طريقة التحليل والتصميم المستخدمة هي Waterfall:

تعريف طريقة الـ Waterfall: هي عملية تصميم متتابعة، وغالبا ما تستخدم في عمليات تطوير النظم والبرمجيات، والتي يعتبر التقدم كالتدفق بشكل مطرد نحو الأسفل (مثل الشلال).



1- **تعريف النطاق (Scope Definition)** :

1.1- تعريف حدود ونص المشكلة:

عدم وجود نظام معلومات لتحليل علامات الطلاب يقوم بـ :

ووضع معايير استنادا إلى علامات الشهادة الثانوية واختبار القبول في المعهد العالي لإدارة الاعمال، ولرفع مستوى اختبار القبول لاختيار طلاب بمستويات أفضل ليتمكنوا من النجاح في مواد السنوات الثلاثة الأولى.

قادر على أرشفة نتائج التحليل، ووضعها في قواعد (Rules) واضحة يمكن الاستفادة منها في وضع معايير لمستوى الطلاب ولاختيار الطلاب لاختصاصاتهم في السنوات الرابعة والخامسة.

2.1- احتياجات قسم إدارة العمليات ونظم المعلومات الإدارية:
تحليل علامات الطلاب المتقدمين لاختبار القبول، وتحليل علامات الطلاب في مرحلة الإجازة التحضيرية في الإدارة.

3.1- المستخدمون:
موظف القسم.

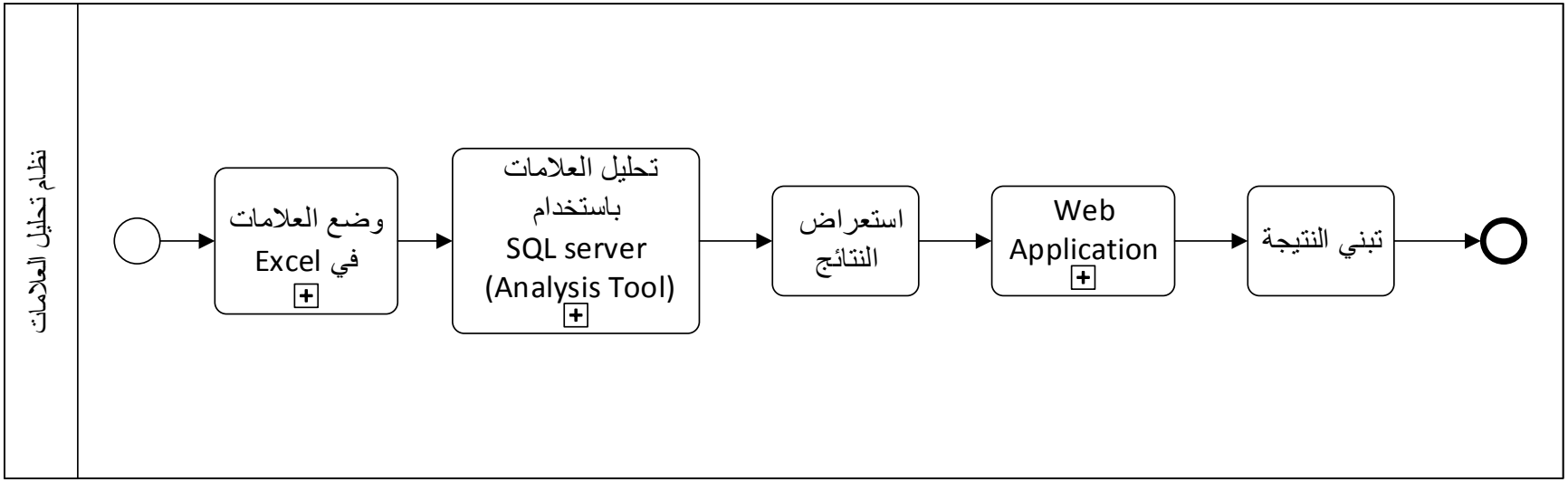
4.1- الهدف:
تحليل العلامات للتوصل إلى قواعد (Rules) ومن خلالها يتم وضع معايير.
نصح الطلاب باختيار اختصاصاتهم، رفع أو خفض مستوى اختبار القبول حسب نتائج التحليل.

2- **تحليل المشكلة (Problem Analysis)** :

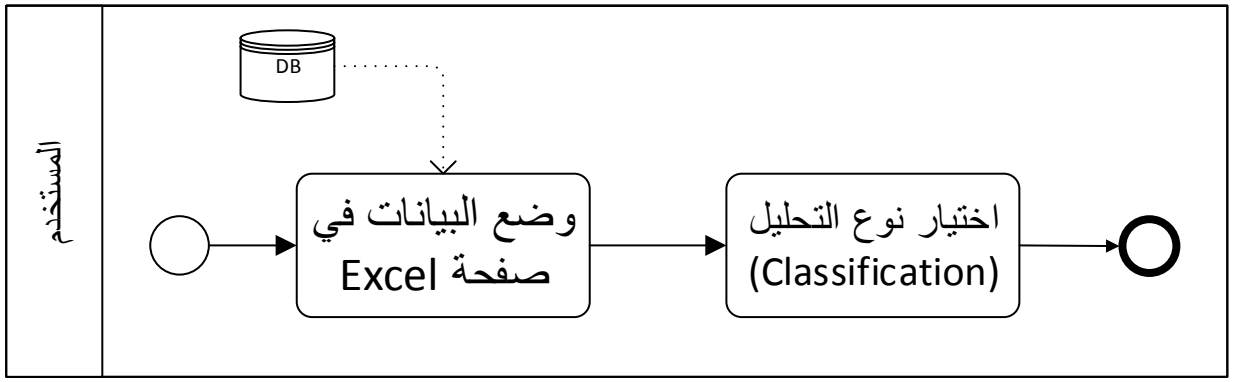
1.2- معرفة مجال المشكلة (Understand The Problem Domain):
كثرة عدد الطلاب ووجود الكثير من التطورات والتعديلات مما يصعب عملية تحليل بيانات الطلاب يدويا.
صعوبة اتخاذ إجراءات (قرارات) بعد التحليل، لعدم وجود معايير يستند إليها.
صعوبة أرشفة نتائج التحليل اليدوي، والاطلاع عليها من قبل أية جهة.

2.2- تحليل المشكلة و الفرص (Analysis Problem and Opportunities):
الحل يكون ببناء نظام معلومات لتحليل علامات الطلاب في المعهد العالي لإدارة الأعمال، يؤمن كافة متطلبات واحتياجات قسم إدارة العمليات ونظم المعلومات الإدارية فيما يخص تحليل علامات الطلاب.

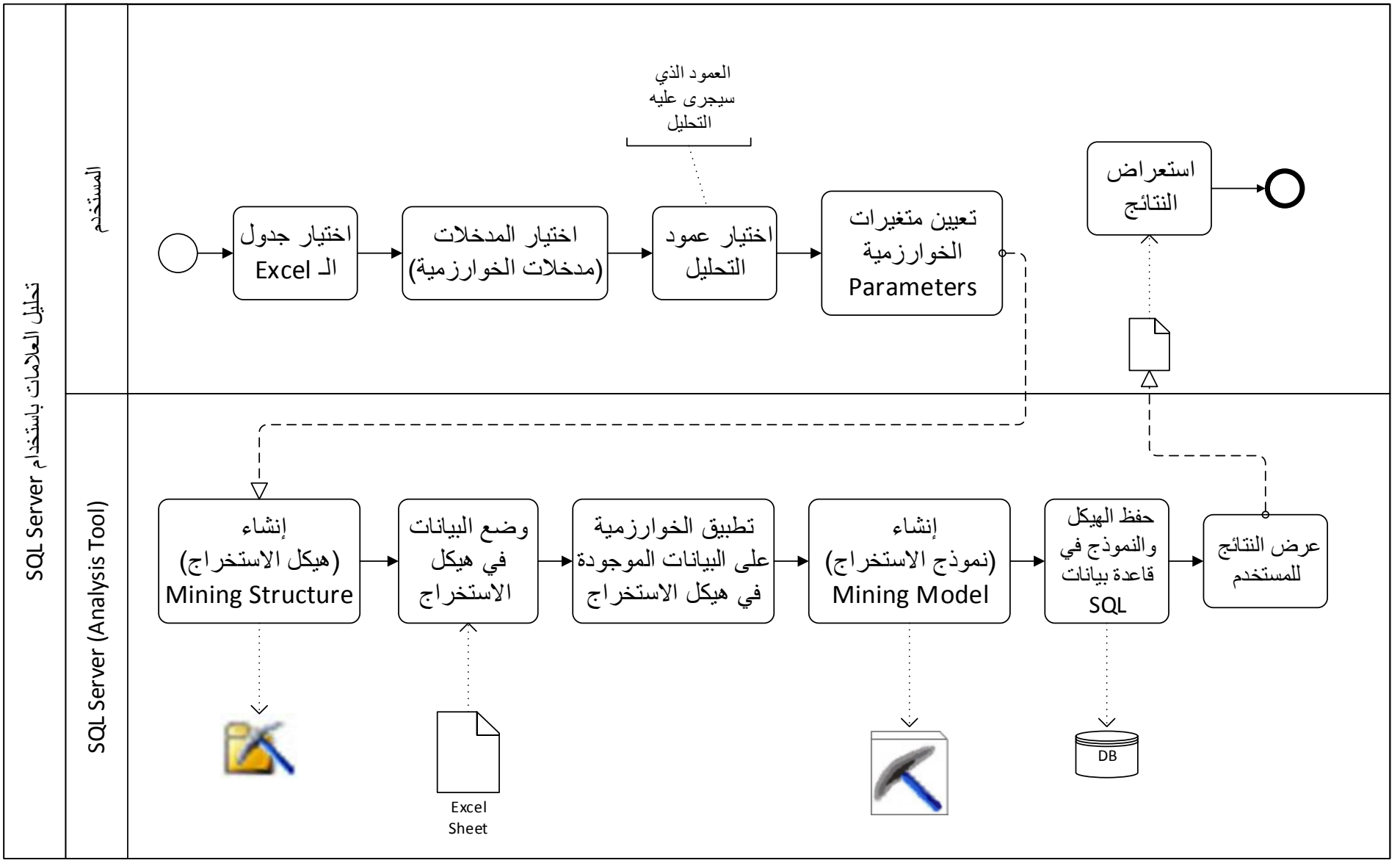
3.2- عمليات الأعمال (Business Process):
باستخدام نموذج BPMN (Business Process Modeling and Notation) لتحليل العمليات، وهذا النموذج يستخدم لتبسيط وفهم مراحل وخطوات العملية (Process) بالإضافة إلى العملية الفيزيائية (Physical Process).


نظام تحليل العلامات
وضع العلامات في Excel
تحليل العلامات باستخدام SQL server (Analysis Tool)
استعراض النتائج
Web Application
تبني النتيجة

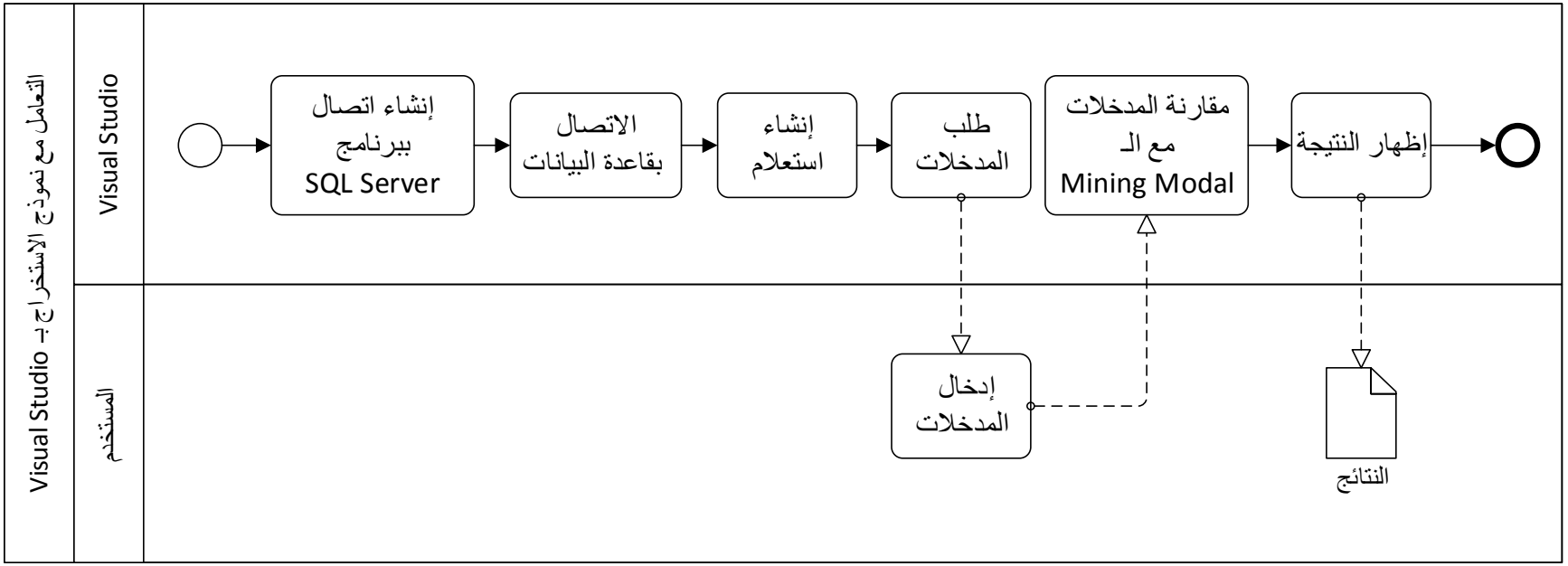
وضع العلامات في Excel:



تحليل العلامات باستخدام SQL Server (Analysis Tool):



التعامل مع نموذج الاستخراج بـ Visual Studio:



3- **تحليل المتطلبات (Requirements analysis):**

تحديد وشرح المتطلبات (Identify and Express Requirements):

1.3- المتطلبات الوظيفية (Functional Requirements):

1.1.3- تخزين علامات الطلاب:

- المدخلات: علامات كل طالب.
- المعالجة: وضع علامات الطلاب في صفحة اكسل.
- المخرجات: أسماء الطلاب وعلاماتهم مرتبة.

2.1.3- تخزين هياكل ونماذج الاستخراج (Mining Structure + Mining Model):

- المدخلات: علامات وأسماء الطلاب.
- المعالجة: تطبيق الخوارزمية على البيانات واستنباط قواعد Rules وحفظهم في قاعدة البيانات.
- المخرجات: Mining Structure + Mining Model، بالإضافة إلى مخططات بيانية (شجرة القرار).

3.1.3- إظهار نتائج المقارنة للمستخدم:

- المدخلات: مدخلات يدخلها المستخدم.
- المعالجة: مقارنة المدخلات مع النماذج المخزنة.
- المخرجات: نتيجة المقارنة.

2.3- متطلبات غير وظيفية (Non Functional Requirements):

1- يجب أن يتم إنجاز العملية التي طلبها المستخدم، مع أمكانية اجراء عدد لا يقل عن 25 عملية بنفس الوقت.

2- يجب أن تكون نسبة حدوث الأخطاء و الأعطال بالنظام منخفضة أو معدومة.

3- الأداء (Performance):

- عدد العمليات والاجرائيات في الثانية : 10 عملية بالثانية.
- زمن الاستجابة للمستخدم للإدخال : 1 ثانية.

4- الموثوقية (Reliability):

- نسبة حدوث الفشل وتكرار الفشل يجب أن لا تزيد عن مرة في السنة.

5- التوفر (Availability):

- احتمال الفشل عن الطلب لا يتجاوز خطأ واحد في كل 100 طلب.

6- الحجم (Size):

- أن لا يتجاوز حجم النظام (ملفات النظام والبرامج الملحقة به)عن الـ 10GB على السيرفر .

7- الاستخدامية (Usability):

- الوقت المستغرق لتعلم استخدام النظام: أن يكون سهل التعلم بالإضافة إلى واجهات سهلة.
- عدد الأخطاء المحدثة من قبل المستخدم: يجب أن يتحمل النظام أخطاء المستخدمين ولا يتأثر بها.

8- قوة البنية (Robustness):

الزمن المستغرق للإعادة إقلاع النظام بعد حدوث فشل، يجب أن لا يتجاوز 10 ثواني إعادة إقلاع وأن يتم الإقلاع أليا.

9- إمكانية الملائمة (Interoperability):

عدد أنظمة التشغيل المستهدفة، إمكانية العمل على هذا النظام على عدد من أنظمة التشغيل (ويندوز 7 – ويندوز xp).

4- **تحليل النظم (System Analysis):**

1.4- الفاعلين (Business Actors):

تعريف الفاعل (Actor): هو أي شيء يحتاج إلى التفاعل مع النظام لتبادل المعلومات، يمكن أن يكون إنسان أو نظام معلومات آخر أو مؤقت (Time Actor) أو حدث يستدعي حدث آخر (Trigger Actor).

- موظف قسم إدارة العمليات ونظم المعلومات الإدارية:

الوصف:

1- وضع العلامات في برنامج الـ Excel.
2- استخدام برنامج SQL Server (Analysis Tool) لتحليل العلامات.
3- استخدام الـ Web-Application لإدخال المدخلات.
4- تصفح النتيجة.

2.4- حالات الاستخدام (Use Case):

تعريف حالات الاستخدام (Use Case): تصف وظائف النظام من وجهة نظر المستخدمين الخارجيين وبالطريقة والمصطلحات التي يفهمونها، وعادة تكون لكل حالة استخدام واجهة للمستخدم.

1- إدخال العلامات.

2- تحليل العلامات.

3- إدخال المدخلات.

4- استعراض النتائج.

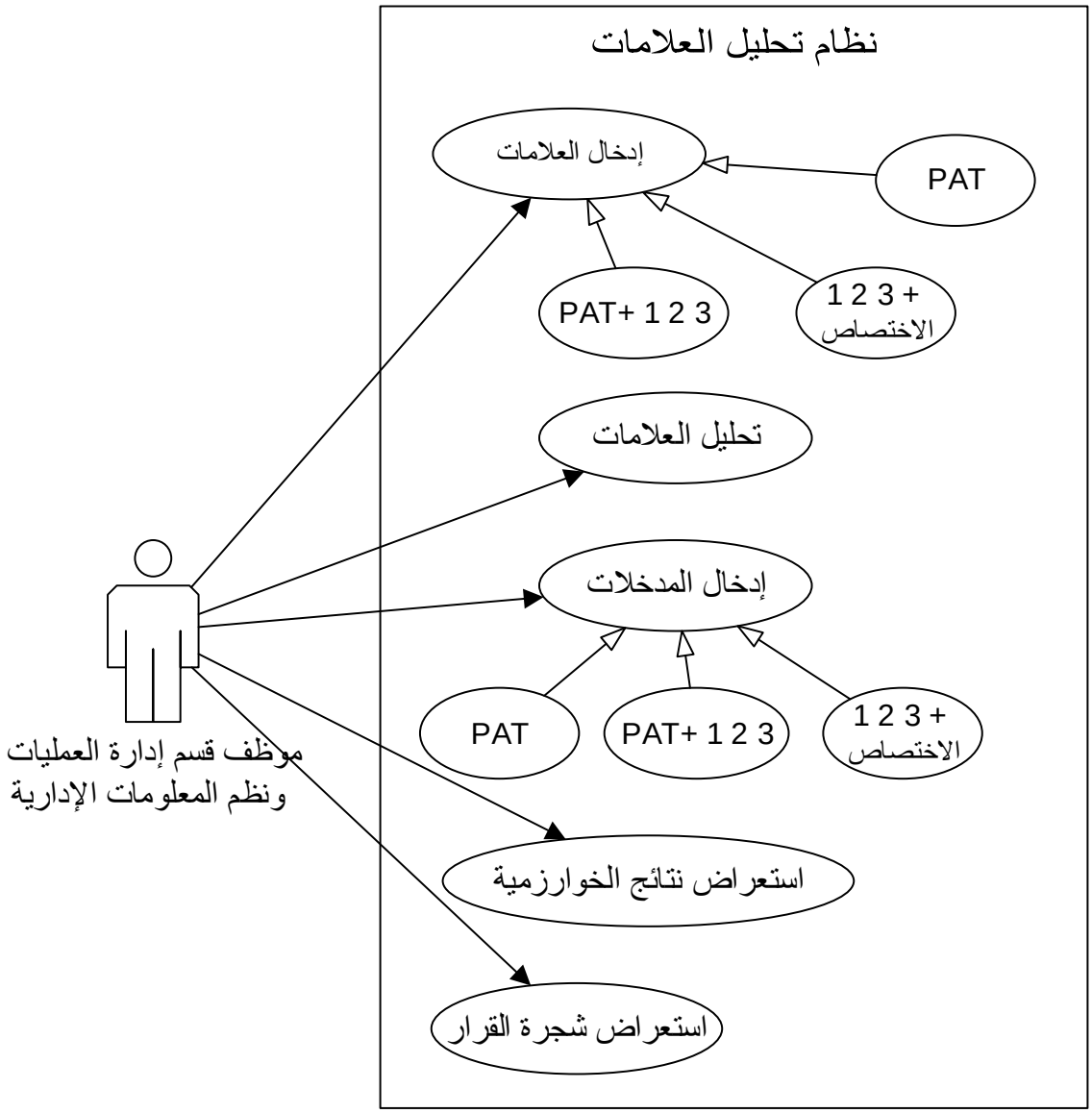
5- استعراض شجرة القرار.

6- حالات استخدام فرعية: (PAT) – (3 2 1 + PAT) – (الاختصاص + 1 2 3).

الرقم	حالة الاستخدام	الوصف	الأدوار والفاعلين والمشاركين
1	إدخال العلامات	هي عملية إدخال العلامات إلى برنامج الـ Excel	موظف قسم إدارة العمليات ونظم المعلومات الإدارية
2	تحليل العلامات	هي عملية تحليل علامات الطلاب باستخدام أداة الـ Analysis Tool	موظف قسم إدارة العمليات ونظم المعلومات الإدارية
3	إدخال المدخلات	هي عملية إدخال المدخلات ليتم مقارنتها مع نموذج الاستخراج (Mining Model)	موظف قسم إدارة العمليات ونظم المعلومات الإدارية
4	استعراض نتائج الخوارزمية	هي عملية استعراض نتائج الخوارزمية	موظف قسم إدارة العمليات ونظم المعلومات الإدارية
5	استعراض شجرة القرار	هي عملية استعراض شجرة القرار الناتجة بعد تطبيق الخوارزمية	موظف قسم إدارة العمليات ونظم المعلومات الإدارية

3.4- مخطط حالات الاستخدام (Use Case Diagram):

تعريف مخطط حالات الاستخدام (Use Case Diagram): هو رسم تخطيطي يصور التفاعلات بين النظام والأنظمة الخارجية والمستخدمين، فهو يصف بيانيا الفاعلين اللذين يستخدمون النظام وبأي طرق يتوقع للمستخدم التفاعل مع النظام.



4.4- الوصف النصي لحالات الاستخدام (Business Requirements Use Case Narrative):

تعريف الوصف النصي لحالات الاستخدام (Business Requirements Use Case Narrative):
هو وصف نصي لحالة الاستخدام وكيفية استدعائها والفاعلين الذين يقومون باستدعائها وكيفية تفاعل الفاعلين مع النظام عن طريقها.

Use-Case Name:	إدخال العلامات	**Use-Case Type:** Business Requirements
Use-Case ID:	1	
Priority:	Normal	
Source:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Primary Business Actors:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Description:	هي عملية إدخال العلامات إلى برنامج الـ Excel، وهذه العمية يقوم بها الفاعل يدويا.	
Precondition:	-	
Trigger:	تبدأ حالة الاستخدام عند إدخال علامات الطلاب إلى برنامج الـ Excel يدويا.	
Typical Course of Events:	**Actor Action**	**System Response**
	1- اختيار نوع علامات التحليل (PAT – 1 2 3 – اختصاص). 2- نسخ العلامات إلى صفحة Excel.	(لا يوجد رد فعل للنظام لأن العملية يدوية)
Alternate Courses:	-	
Conclusion:	عند انتهاء الفاعل من العملية يكون قد أدخل العلامات إلى برنامج الـ Excel.	
Business Rules:	أول خطوة في عملية التحليل هي وضع العلامات في برنامج الـ Excel.	

Use-Case Name:	تحليل العلامات	**Use-Case Type:** Business Requirements
Use-Case ID:	2	
Priority:	Normal	
Source:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Primary Business Actors:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Description:	هي عملية تحليل علامات الطلاب باستخدام أداة الـ Analysis Tool الموجودة في برنامج SQL Server، يدويا.	
Precondition:	أن تكون البيانات جاهزة وموجودة في صفحة Excel.	
Trigger:	تبدأ حالة الاستخدام عند فتح أداة الـ Analysis Tool للبدء بتحليل العلامات.	
Typical Course of Events:	**Actor Action**	**System Response**
	1- تحديد العلامات بصفحة الـ Excel. 2- اختيار الخوارزمية المناسبة (Classification). 3- اختيار مدخلات (مدخلات الخوارزمية). 4- اختيار عمود التحليل. 5- تعيين متغيرات الخوارزمية.	6- إنشاء Mining Structure. 7- إنشاء Mining Model. 8- حفظ الـ MS + MM في قاعدة البيانات SQL.
Alternate Courses:	-	
Conclusion:	عند انتهاء الفاعل من العملية يكون قد قام بتحليل البيانات.	
Business Rules:	ثاني خطوة في عملية التحليل هي تحليل البيانات باستخدام الـ Analysis Tool.	

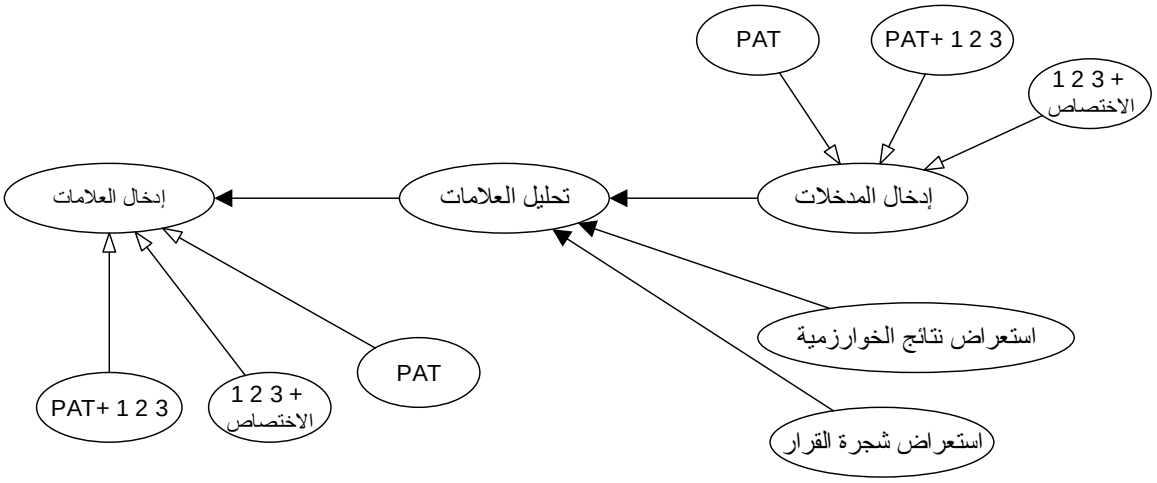
Use-Case Name:	إدخال المدخلات	**Use-Case Type:** Business Requirements
Use-Case ID:	3	
Priority:	Normal	
Source:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Primary Business Actors:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Description:	هي عملية إدخال المدخلات ليتم مقارنتها مع نموذج الاستخراج (Mining Model) الناتجة عن تحليل البيانات.	
Precondition:	أن تكون نماذج الاستخراج (Mining Model) منشأة ومحفوظة في قاعدة البيانات.	
Trigger:	تبدأ حالة الاستخدام عند فتح صفحة الويب الخاصة بالنظام واختيار النوع: (PAT) – (PAT + 3 2 1) – (3 2 1 + اختصاص).	
Typical Course of Events:	**Actor Action**	**System Response**
	1- طلب صفحة الويب الخاصة بالنظام.	
		2- فتح صفحة الويب التي تحتوي أنواع التحليل.
	3- تحديد نوع التحليل.	
		4- فتح صفحة الويب التي تحوي نوع التحليل المختار.
	5- إدخال المدخلات.	
	6- الضغط على موافق.	
		7- إنشاء اتصال مع برنامج SQL-Server.
		8- إنشاء اتصال مع قاعدة البيانات.
		9- إنشاء استعلام DMX.
		10- مقارنة المدخلات مع نموذج الاستخراج Mining Model.
	12- تصفح النتيجة.	11- إظهار النتيجة.
Alternate Courses:	2- 4- 7- 8- 9- 10- 11- إظهار رسالة وجود خطأ.	
Conclusion:	عند انتهاء الفاعل من العملية يكون قد قام بإدخال البيانات لمقارنتها.	
Business Rules:	ثالث خطوة في التحليل هي إدخال المدخلات الجديدة لمقارنتها.	

Use-Case Name:	استعراض نتائج الخوارزمية	**Use-Case Type:** Business Requirements
Use-Case ID:	4	
Priority:	Normal	
Source:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Primary Business Actors:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Description:	هي عملية استعراض نتائج الخوارزمية، يدويا.	
Precondition:	أن تكون البيانات قد أدخلت إلى صفحة Excel، وتم تحليلها باستخدام أداة الـ Analysis Tool.	
Trigger:	تبدأ حالة الاستخدام عند فتح نموذج الاستخراج Model Mining المخزن في قاعدة بيانات التحليل لبرنامج SQL Server (Analysis Tool) .	
Typical Course of Events:	**Actor Action**	**System Response**
	1- فتح برنامج SQL Server. 2- الاتصال بقاعدة البيانات. 3- اختيار هيكل استخراج Mining structure. 4- استعراض النموذج Mining Model. 5- تصفح نتائج الخوارزمية.	(لا يوجد رد فعل للنظام لأن العملية يدوية)
Alternate Courses:	-	
Conclusion:	عند انتهاء الفاعل من العملية يكون قد قام باستعراض نتائج الخوارزمية.	
Business Rules:	هذه الخطوة تفيد في استعراض نتائج الخوارزمية لبناء المعايير استنادا إلى النتائج.	

Use-Case Name:	استعراض شجرة القرار	**Use-Case Type:** Business Requirements
Use-Case ID:	5	
Priority:	Normal	
Source:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Primary Business Actors:	موظف قسم إدارة العمليات ونظم المعلومات الإدارية	
Description:	هي عملية استعراض شجرة القرار، يدويا.	
Precondition:	أن تكون البيانات قد أدخلت إلى صفحة Excel، وتم تحليلها باستخدام أداة الـ Analysis Tool.	
Trigger:	تبدأ حالة الاستخدام عند فتح نموذج الاستخراج Mining Model المخزن في قاعدة بيانات التحليل لبرنامج SQL Server (Analysis Tool) .	
Typical Course of Events:	**Actor Action**	**System Response**
	1- فتح برنامج SQL Server. 2- الاتصال بقاعدة البيانات. 3- اختيار هيكل استخراج Mining structure. 4- استعراض النموذج Mining Model. 5- تصفح شجرة القرار.	(لا يوجد رد فعل للنظام لأن العملية يدوية)
Alternate Courses:	-	
Conclusion:	عند انتهاء الفاعل من العملية يكون قد قام بتصفح شجرة القرار.	
Business Rules:	هذه الخطوة تفيد في استعراض شجرة القرار لبناء المعايير استنادا إلى النتائج.	

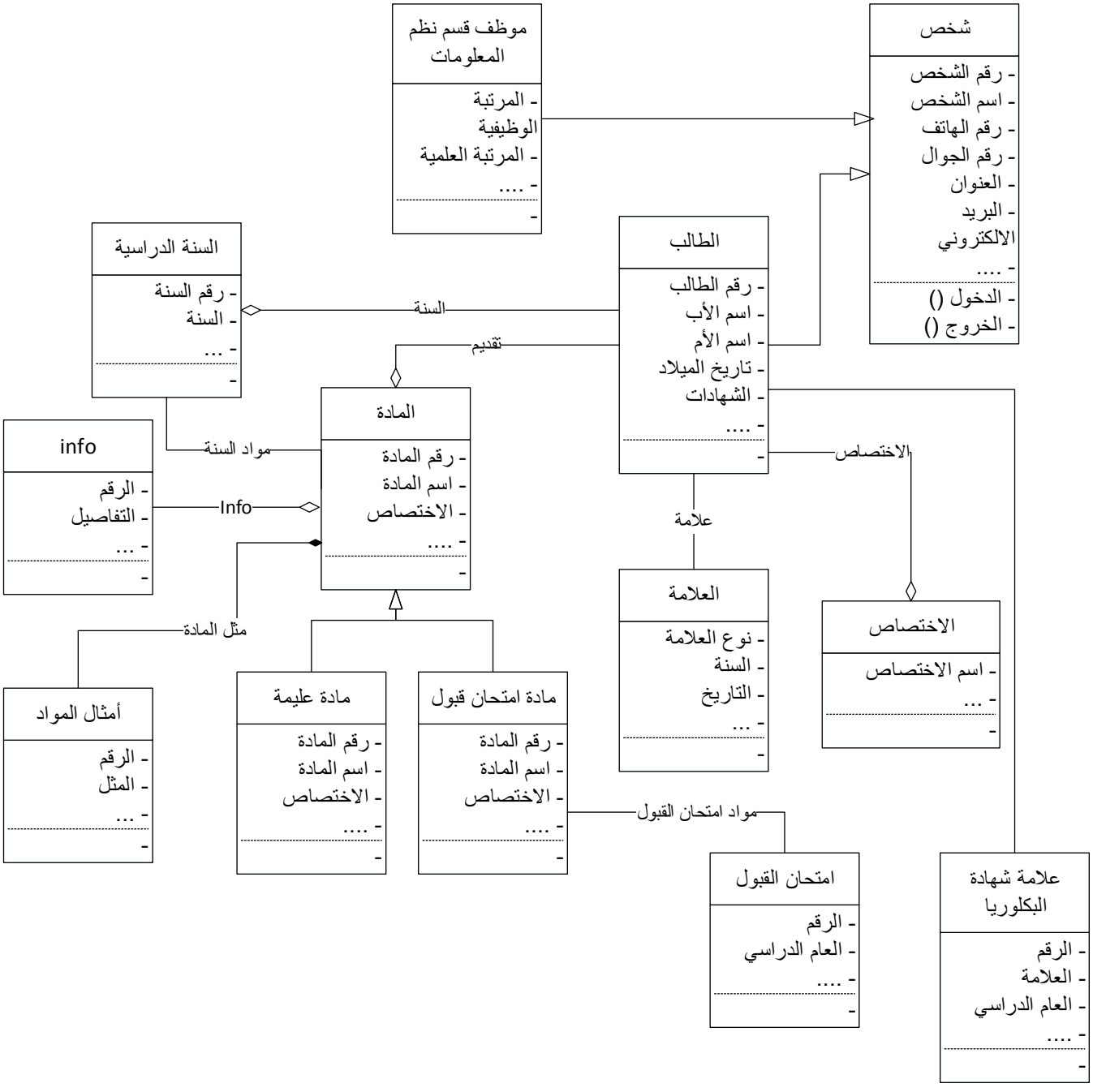
5.4- مخطط التبعية (Dependency Diagram):

تعريف مخطط التبعية (Dependency Diagram): هو رسم وصفي للتبعيات بين حالات الاستخدام، حيث يظهر ترتيب حالات الاستخدام وتبعيتها.



6.4- مخطط الصفوف Class Diagram:

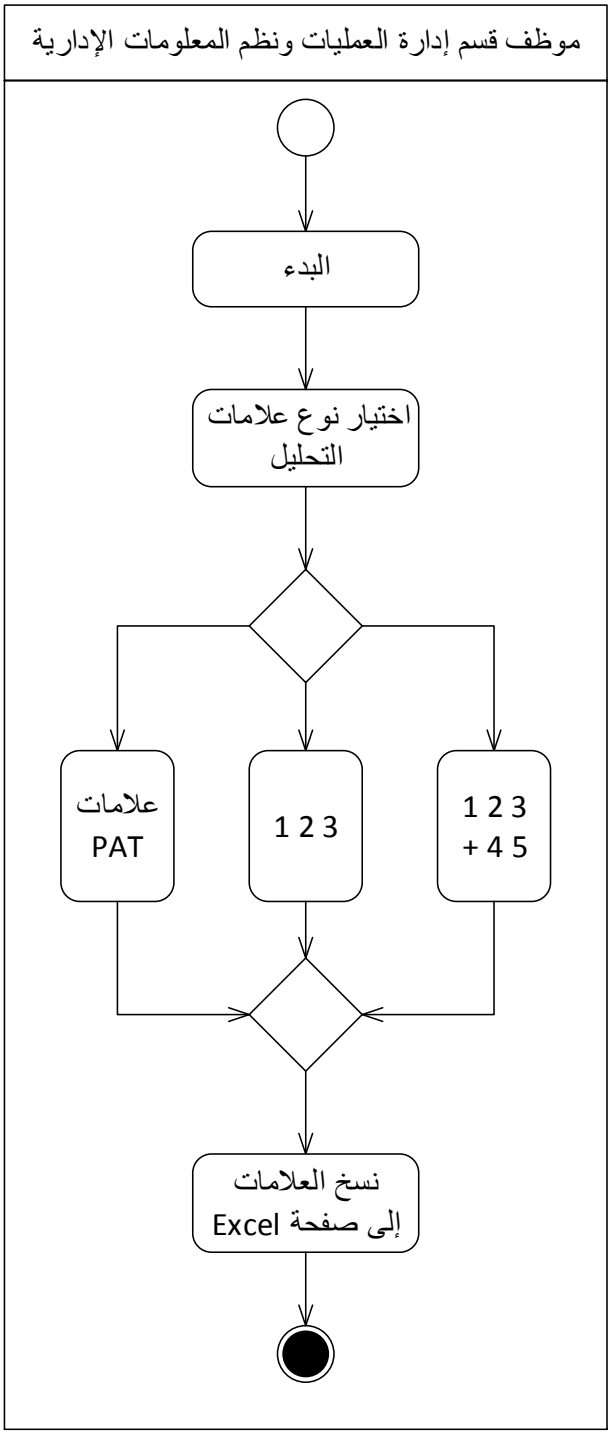
تعريف مخطط الصفوف (Class Diagram): هو رسم وصفي لصفوف وكائنات النظام، ويظهر طبقات الكائن والعلاقات بين الصفوف.



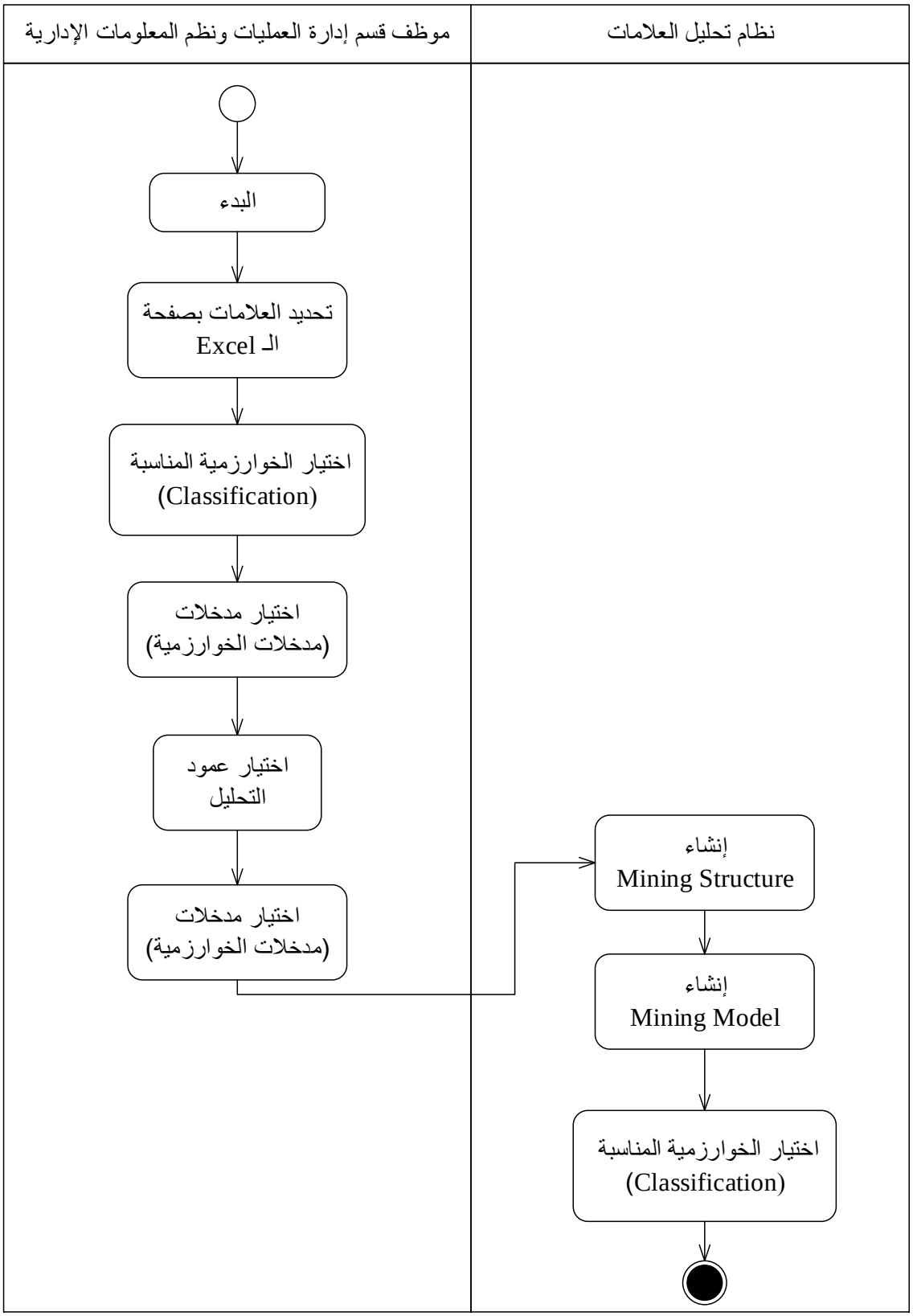
7.4- مخطط النشاط (Activity Diagram):

تعريف مخطط النشاط (Activity Diagram): هو رسم تخطيطي يمكن استخدامه لتصوير تدفق العملية، ويمكن رسمه من خلال بند (**Typical Course of Events**:) الموجود في الوصف النصي لحالة الاستخدام.

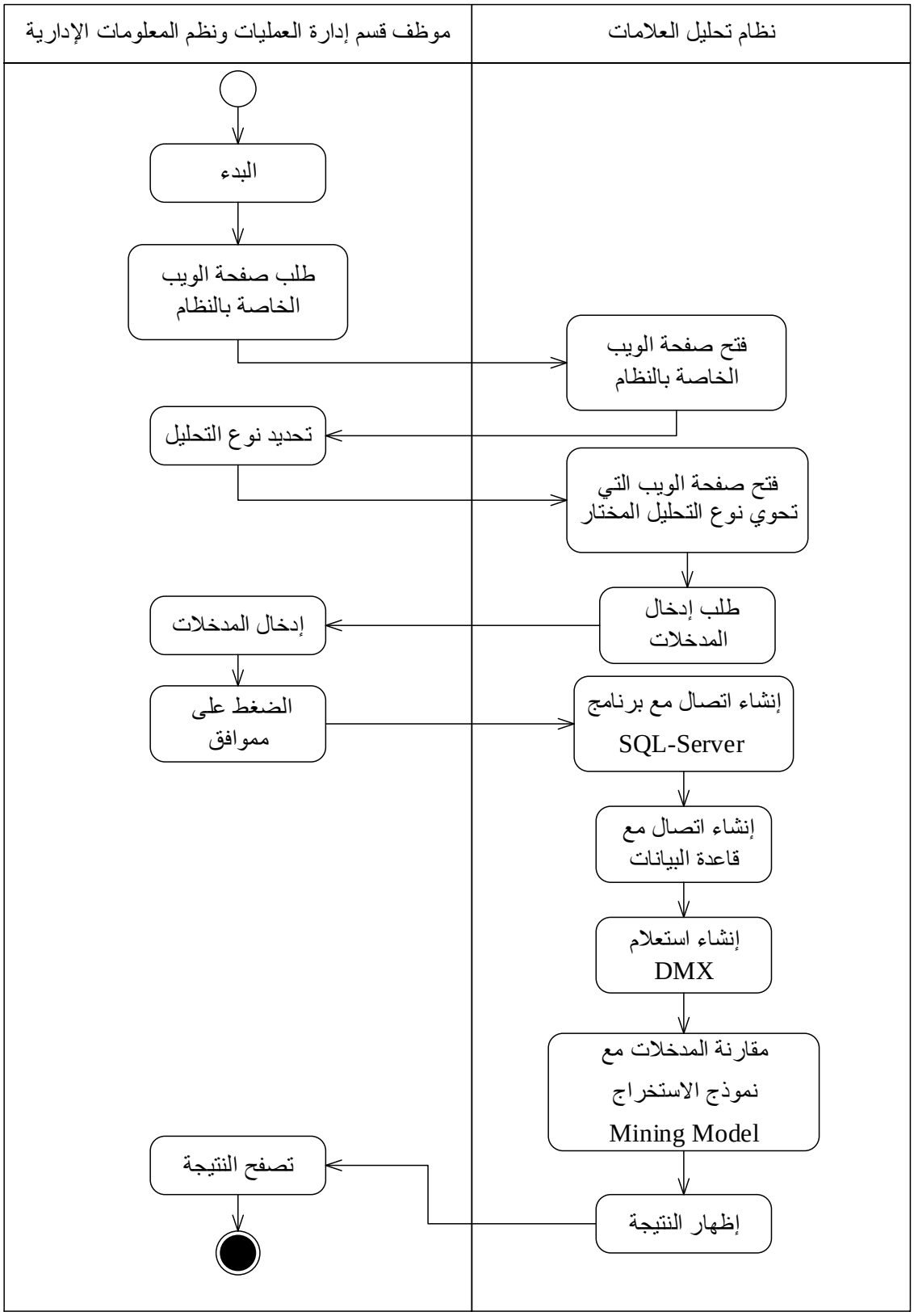
- حالة استخدام (إدخال العلامات):



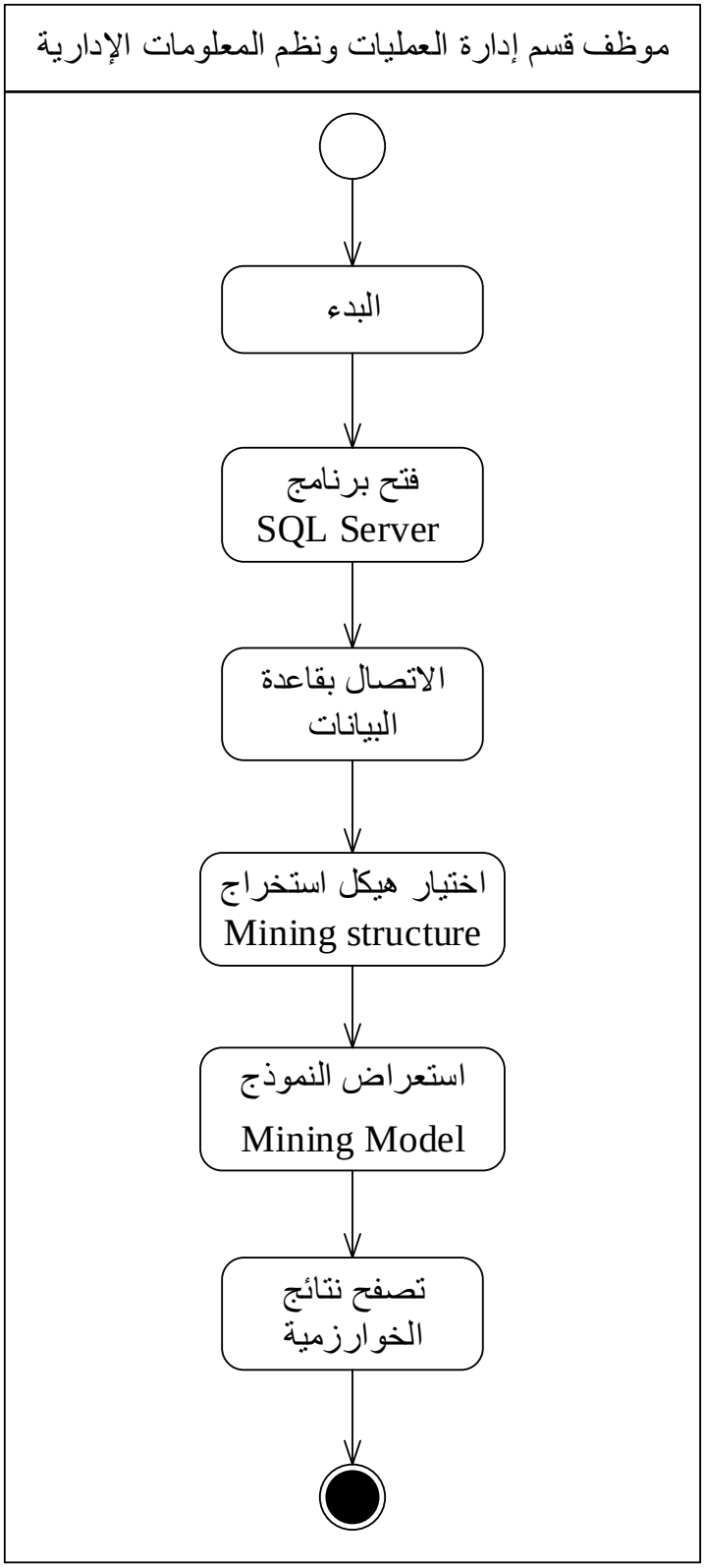
- حالة استخدام (تحليل العلامات):



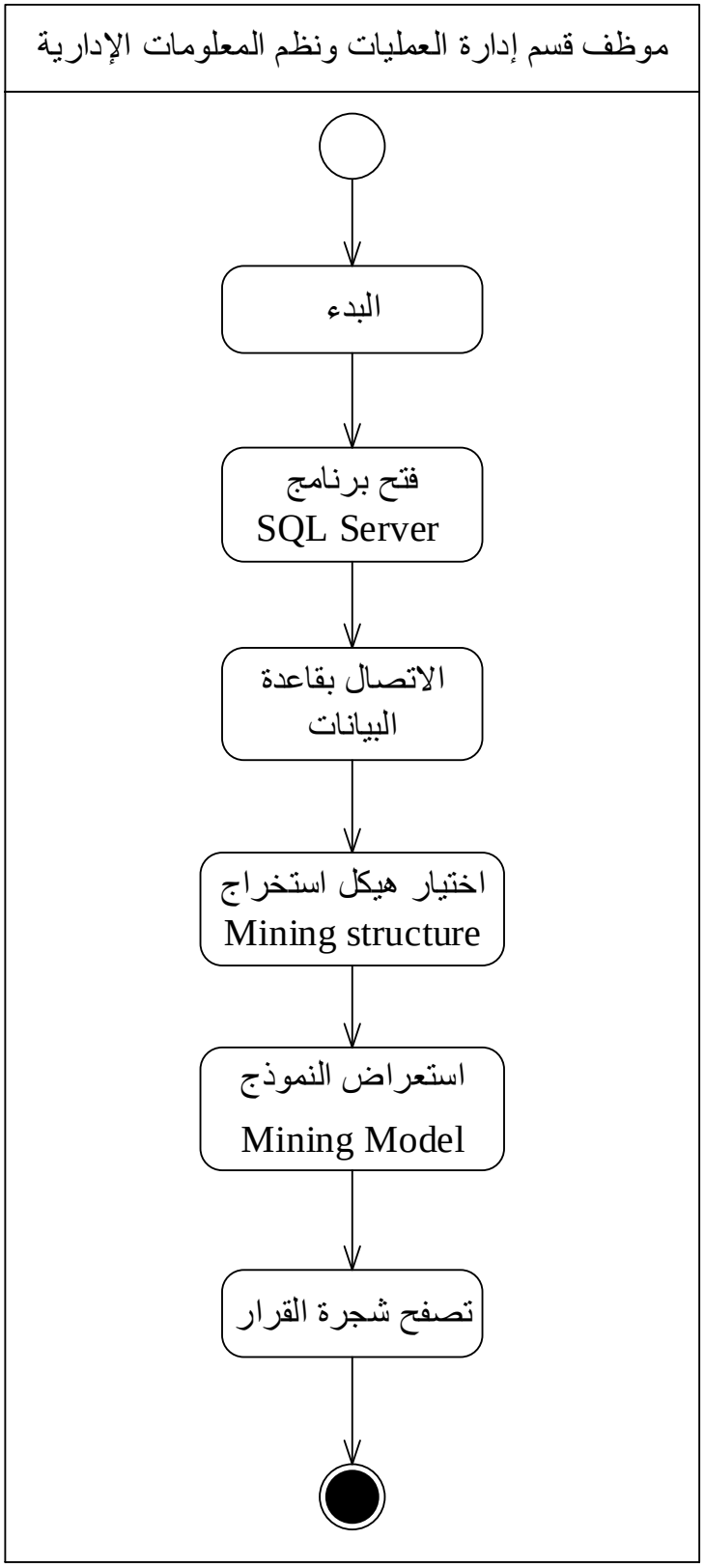
- حالة استخدام (إدخال المدخلات):



- حالة استخدام (استعراض نتائج الخوارزمية):



- حالة استخدام (استعراض شجرة القرار):



5- تصميم النظم (System Design) :

سيتم استخدام نظام محلي (local system) حيث تتم فيه جميع العمليات والفعاليات في النظام، حيث يقوم الفاعل بالعمل على حاسبه.

1.5- التكنولوجيا المستخدمة في النظام: (التقرير الفني) :

1- برنامج الـ Excel الموجود في حزمة برمجيات الـ Office.

2- تطبيق Data Mining Add-Ins for Office 2010.

3- SQL Server 2008.

4- Web-Application (build with Visual Studio 2012).

1.1.5- برنامج Excel:

يُعتبر برنامج Excel من البرامج المكتبية المتعدد الأغراض، مصمم لإجراء العمليات المتنوعة عن طريق خلايا (Cells) يتم قراءة محتوياتها إلكترونياً.

والجدول الإلكتروني (Spread sheet) هو جدول مقسم إلى صفوف (Rows) وأعمدة (Columns).

نحتاج إلى البرنامج في عملية التحليل لتخزين البيانات بطريقة سهلة للقيام بعملية تحليلها لاحقا.

2.1.5- Data Mining Add-Ins for Office 2010:

تطبيق وسيط بين برنامج SQL Server 2008 وبرنامج Excel، حيث يمكننا من خلال هذه الأداة أن نقوم بالتحكم بالأداة Analysis Tool الموجودة في SQL Server، للقيام بعملية الاستخراج من البيانات (Data Mining).

حيث يمكننا اختيار الخوارزمية وإدخال البيانات من صفحة Excel دون الحاجة للدخول لبرنامج SQL Server، وتظهر النتائج ومخطط الخوارزمية في نافذة تظهرها الأداة، ويمكن حفظ هذه النتائج والمخطط ضمن صفحة Excel.

3.1.5- SQL Server 2008 (Analysis Tool):

هو برنامج لإدارة قواعد البيانات من شركة مايكرو سوفت، ومن مهامه حفظ البيانات وترتيبها واستحضارها وتعديلها والإضافة عليها.

اللغة التي يتعامل بها البرنامج هي (Structure Query Language) SQL.

ولهذا البرنامج العديد من الخدمات (Services) والأدوات (Tools) ومن أهم هذه الأدوات هي Analysis Tool التي سنستخدمها في عملية تحليل العلامات.

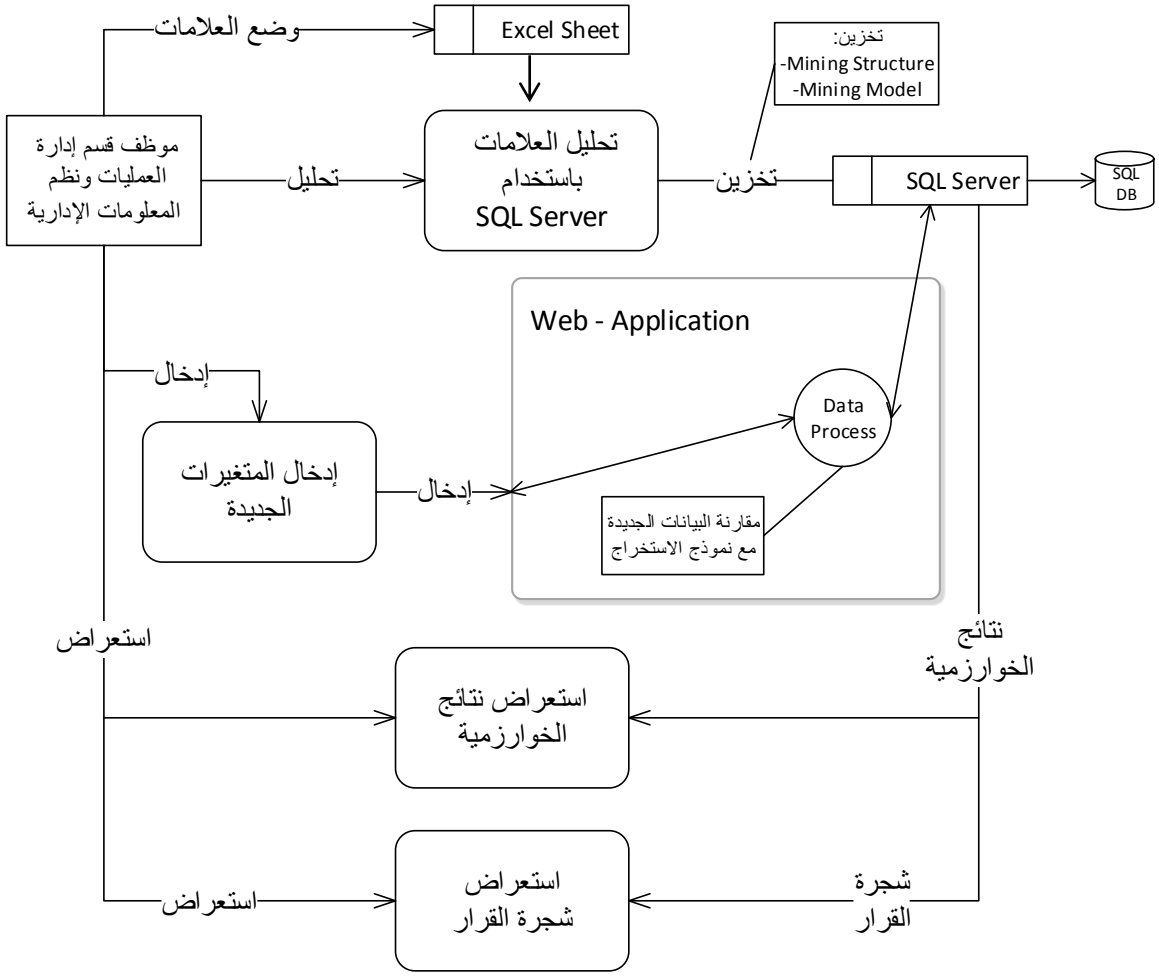
حيث تستخدم هذه الأداة في تطبيق خوارزميات مختلفة على البيانات لتحليلها واستخراج المعارف من هذه البيانات.

4.1.5- Web-Application (build with Visual Studio 2012):

واجهات ويب مبرمجة باستخدام برنامج Visual Studio 2012 تستخدم لإدخال بيانات جديدة ومقارنتها مع نموذج الاستخراج (Mining Model) الذي تم إنشاؤه وحفظه في قاعدة بيانات SQL Server، عن طريق إنشاء استعلام لمقارنة البيانات الجديدة والمحفوظة.

2.5- المخطط الفيزيائي لتدفق البيانات (Physical Data Flow Diagram) PDFD:

هو مخطط يمثل تدفق البيانات في نظام المعلومات ويوضح جهة تدفقها بين الفاعلين والنظام وقاعدة بيانات النظام، ويسهل عملية فهم تدفق البيانات.



3.5- تصميم نظام تحليل العلامات:

نظام تحليل العلامات يتألف من ثلاثة كتل من البرمجيات لوجود ثلاثة أنواع من التحليل كما ذكرنا سابقا:

(تحليل علامات الـ PAT ، تحليل علامات الـ PAT + علامات السنوات 1 2 3 ،

تحليل علامات السنوات 1 2 3 + 4 5 لمعرفة الاختصاصات).

1- نظام تحليل علامات الـ PAT :

نقوم بتحليل علامات الـ PAT لاستنباط قواعد من العلامات وهذه القواعد تفيدنا في وضع معايير لقبول الطلاب أو معايير لمستوى الطلاب أو معايير لمستوى اختبار القبول.
بالإضافة إلى إدخال قيم جديدة والحصول على النتائج (مقارنة البيانات الجديدة مع القواعد الناتجة عن الخوارزمية).

1.1- التحليل:

1- تحليل (معدل البكلوريا) والنجاح في اختبار القبول (Info).
2- تحليل (مادة الرياضيات والمعلوماتية) والنجاح في اختبار القبول (Info).
3- تحليل (مادة الانكليزي) والنجاح في اختبار القبول (Info).
4- تحليل (مادة الإدارة) والنجاح في اختبار القبول (Info).
5- تحليل مواد امتحان القبول (الرياضيات والمعلوماتية – الإنكليزي – الإدارة) والنجاح في اختبار القبول (Info).
6- تحليل المعدل (الامتحان – معدل البكلوريا) والنجاح في اختبار القبول (Info).
7- تحليل (معدل البكلوريا – رياضيات ومعلوماتية – إنكليزي – إدارة – معدل) والنجاح في اختبار القبول (info).

2.1- سنأخذ تحليل (مادة الرياضيات والمعلوماتية) مثالا وبقية أنواع التحليل مذكورة بشكل مفصل في الملحق الموجود في الـ CD المرفق.

- الخطوات:

1.2.1- إدخال البيانات لصفحة Excel:

واصفات البيانات في حالة تحليل علامات الـ PAT:

M	L	K	J	I	H	G	F	E	D	C	B	A	
info	الحالة	معدل	محصل	إدارة	إنكليزي	رياضيات ومعلوماتية	M_B	معدل البكلوريا	العام	الكنية	الاسم	الرقم	1
Yes	مقبول	J2+F2	I2+H2+G2				round(E2,2)					1	2
No	مرفوض												3
													4

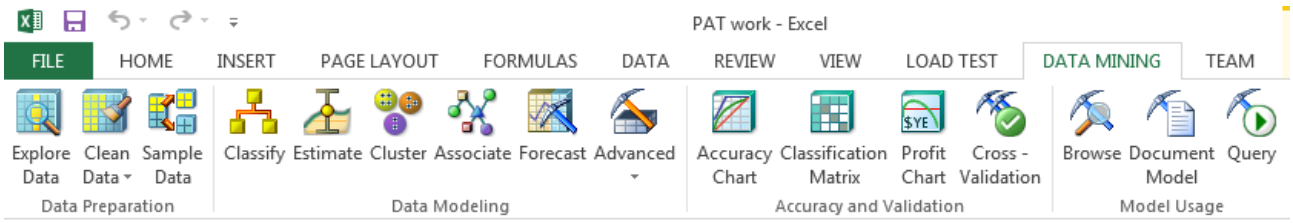
مع التنويه أن عمود info يحوي:Yes ناجح في امتحان القبول (مقبول)، No راسب في امتحان القبول (مرفوض).

2.2.1- تحليل البيانات باستخدام برنامج (Analysis Tool) SQL Server 2008:

سنأخذ مثال: تحليل (مادة الرياضيات والمعلوماتية) والنجاح في امتحان القبول (Info):

أولا- تحديد جدول العلامات الموجود في صفحة الـ Excel.

ثانيا- اختيار خيار Data Mining في شريط الخيارات في Excel:

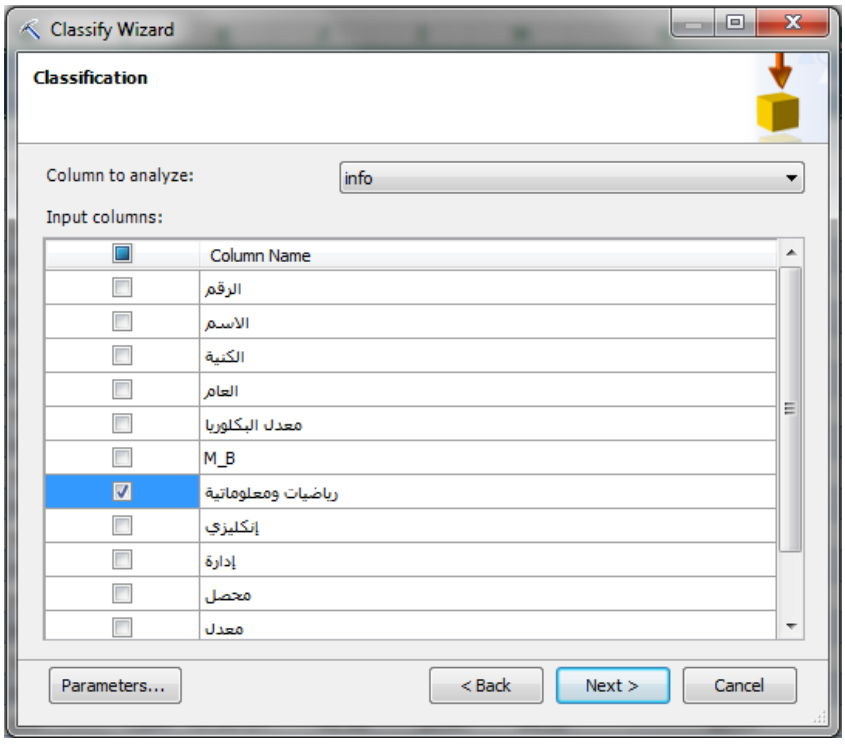


ثالثا- اختيار نموذج استخراج البيانات (Data Mining Model):

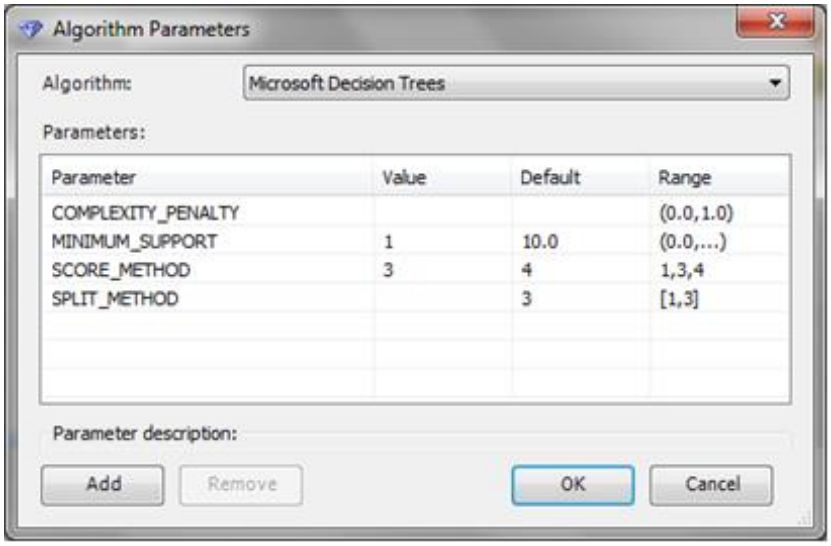
النموذج المستخدم في نظام تحليل العلامات هو Classification التصنيف، لتصنيف علامات الطلاب ضمن قواعد (مثال: علامة البكلوريا > 55 = yes) وهذه القواعد تفيد في وضع معايير.

رابعا- تحديد مدخلات الخوارزمية (input columns) والعمود الذي سيتم عليه التحليل (Column to Analyze):

حيث اخترنا مدخلات الخوارزمية (رياضيات ومعلوماتية) والعمود المراد تحليله (info)



خامسا– نحدد الخوارزمية ومتغيرات الخوارزمية من الخيار (Parameters):

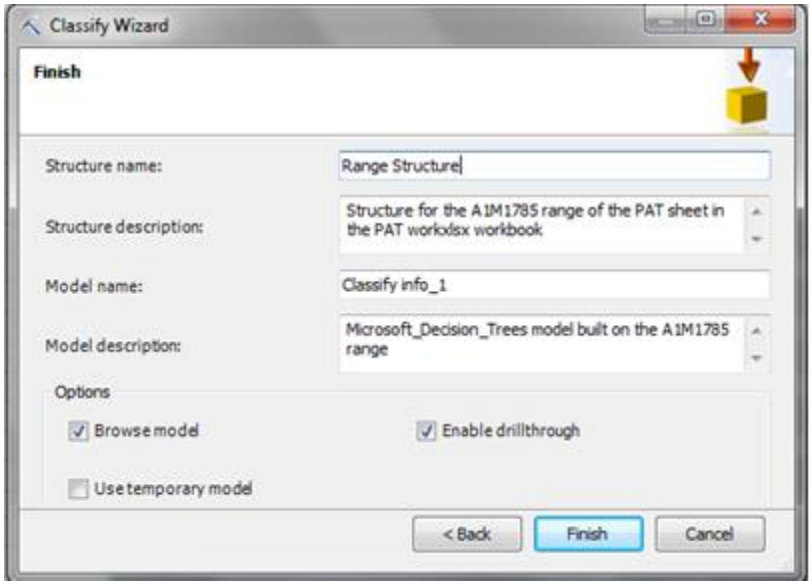


نختار الـ Minimum_Support = 1 لأننا نريد أخذ كافة العناصر لتحليلها.

و نختار طريقة الحساب الـ Score_Method = 3 طريقة نظرية Bayes.

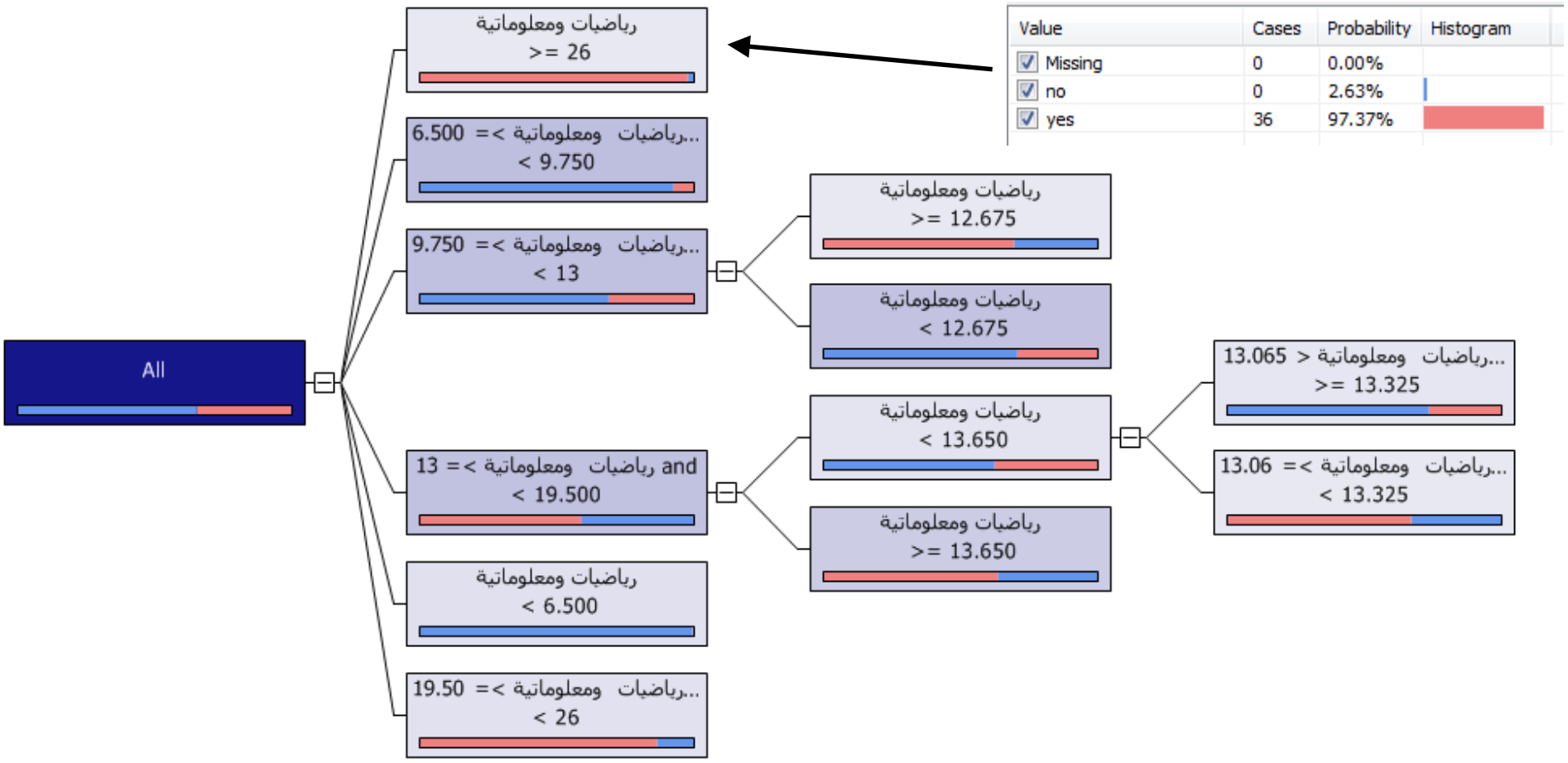
ونختار نوع الخوارزمية المناسبة: ID3 Decision Tree.

سادسا– كتابة اسم هيكل الاستخراج (Mining Structure) و اسم نموذج الاستخراج (Mining Model):



سابعا– بعدها يقوم برنامج SQL Server بإدخال البيانات إلى الخوارزمية وحفظ النتائج في قاعدة بيانات SQL وإظهار نافذة النتائج، القواعد الناتجة عن الخوارزمية وشجرة القرار:

إظهار نافذة النتائج:


All
رياضيات ومعلوماتية
>= 26
6.500 =< رياضيات ومعلوماتية...
< 9.750
9.750 =< رياضيات ومعلوماتية...
< 13
13 =< رياضيات ومعلوماتية and
< 19.500
رياضيات ومعلوماتية
< 6.500
19.50 =< رياضيات ومعلوماتية...
< 26
رياضيات ومعلوماتية
>= 12.675
رياضيات ومعلوماتية
< 12.675
رياضيات ومعلوماتية
< 13.650
رياضيات ومعلوماتية
>= 13.650
13.065 > رياضيات ومعلوماتية...
>= 13.325
13.06 =< رياضيات ومعلوماتية...
< 13.325
Value
Cases
Probability
Histogram
Missing
0
0.00%
no
0
2.63%
yes
36
97.37%

من خلال النتائج الناتجة عن الخوارزمية وشجرة القرار نقوم بكتابة القواعد:

(نسمي مادة رياضيات ومعلوماتية x)

Yes		x	26 <
Yes	< 26	x	19.5 <
?	< 19.5	x	13 <
No	< 13.6	x	13 <
Yes	< 19.5	x	13.6 <
?	< 13	x	9.7 <
Yes	< 13	x	12 <
No	< 12	x	9.7 <
No	< 9.7	x	6.5 <
No	< 6.5	x	

نستنتج من المجالات السابقة القواعد:

Yes		x	13.6 <
Yes	< 13	x	12 <
No	< 13.6	x	13 <
No	< 12	x	9.7 <

القواعد:

(علامة الرياضيات والمعلوماتية > 13.6 = Yes) يقبل الطالب إذا حصل على 13.6 فما فوق.

(علامة الرياضيات والمعلوماتية < 13.6 = No) يرفض الطالب إذا حصل على علامة أقل من 13.6.

3.2.1- استخدام واجهات إدخال Web-Application:

يتم استخدام واجهات إدخال ويب مصممة ببرنامج Visual Studio 2012 لإدخال البيانات الجديدة ومقارنتها مع نموذج الاستخراج (Mining Model) المحفوظ في قاعدة البيانات لإعطاء التنبؤ للبيانات الجديدة.

* الخطوات:

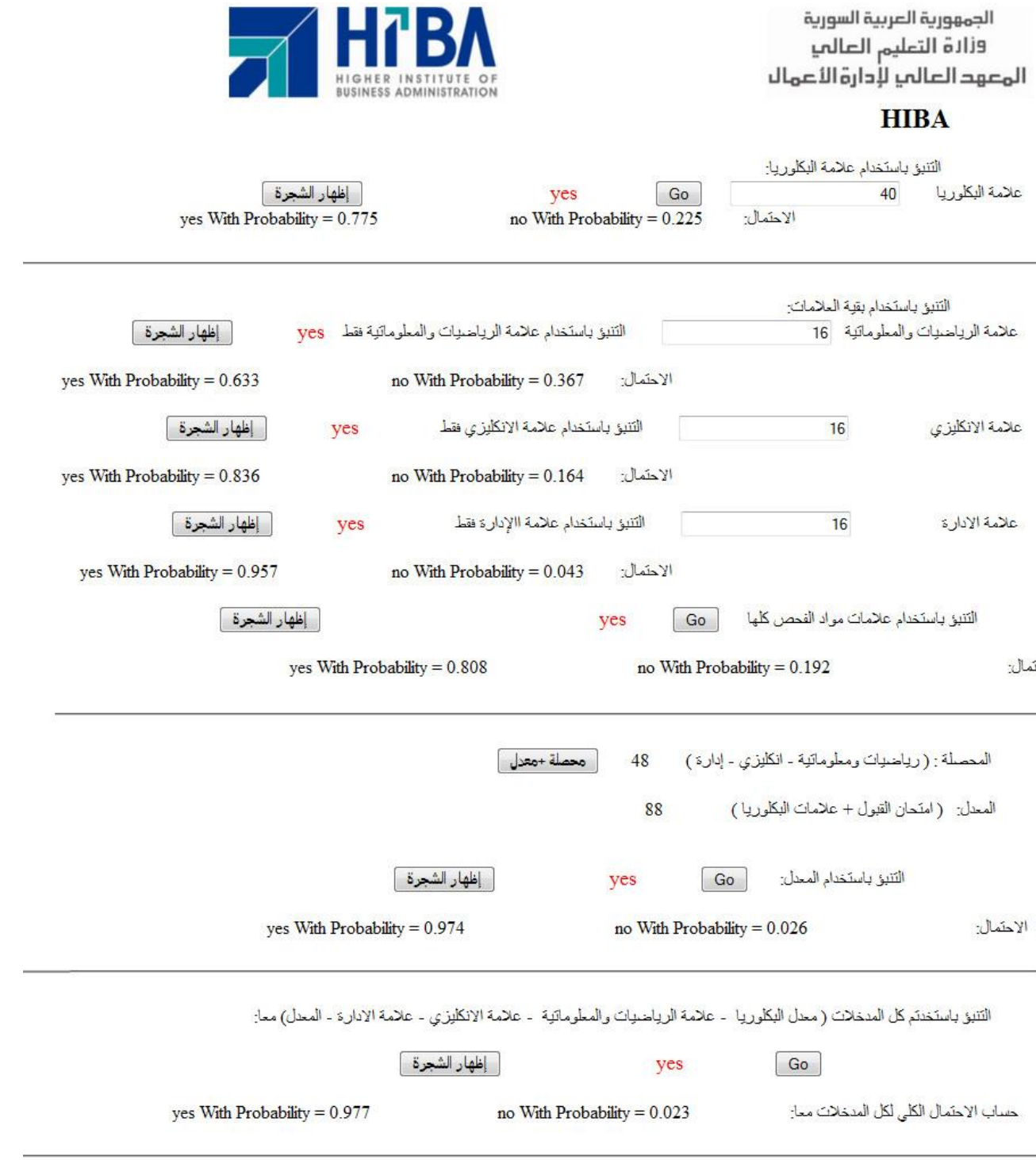
أولا- عند إدخال البيانات الجديدة والضغط على زر (Go) يتم انشاء اتصال ببرنامج SQL Server وانشاء اتصال بقاعدة بيانات SQL.

ثانيا – إنشاء استعلام بالبيانات المدخلة للتنبؤ من نموذج الاستخراج (Mining Model) المنشأ في عملية التحليل.

ثالثا- عرض النتيجة في مربع النص (Yes أو No).

رابعا- عرض احتمال النتيجة في مربع النص (احتمال الـ Yes و الـ No معا).

- واجهة البرنامج:



من خلال العلامات المدخلة والنتائج الظاهرة نجد أن الطالب ناجح باختبار الـ PAT.

2- **نظام تحليل علامات PAT + 1 2 3 :**

نقوم بتحليل علامات مواد اختبار الـ PAT وتأثيره على مواد السنوات الثلاثة الأولى لاستنباط قواعد من العلامات وهذه القواعد تفيدنا في وضع معايير لقبول الطلاب ومعايير لمستوى الطلاب ومعايير لمستوى اختبار القبول ومعايير لمستوى فحص المواد في السنوات الثلاثة الأولى.
بالإضافة إلى إدخال قيم جديدة والحصول على النتائج (مقارنة البيانات الجديدة مع القواعد الناتجة عن الخوارزمية).

1.2- التحليل:

1- تحليل (مادة الـ E في الـ PAT) والعلامات في مواد الـ E في السنوات الثلاثة الأولى (إنكليزي 1 – إنكليزي 2 – إنكليزي 3 – إنكليزي 4).
2- تحليل (مادة المعلوماتية في الـ PAT) والعلامات في مواد المعلوماتية في السنوات الثلاثة الأولى (تطبيقات المعلوماتية – أساسيات البرمجة – قواعد البيانات).
3- تحليل (مادة الرياضيات في الـ PAT) والعلامات في مواد الرياضيات في السنوات الثلاثة الأولى (رياضيات – إحصاء واحتمالات – رياضيات تطبيقية في الإدارة – أساليب البحث العلمي – بحوث العمليات – ...).
4- تحليل (مادة الإدارة في الـ PAT) والعلامات في مواد الاقتصاد والتجارة في السنوات الثلاثة الأولى (مدخل إلى علم الاقتصاد – اقتصاد كلي – اقتصاد إسلامي – تجارة دولية).
5- تحليل (مادة الإدارة في الـ PAT) والعلامات في مواد الإدارة والموارد في السنوات الثلاثة الأولى (مبادئ ووظائف الإدارة – نظريات الإدارة – تنظيم إداري – إدارة الموارد البشرية – سلوك تنظيمي – مسؤولية اجتماعية – إدارة التفاوض).
6- تحليل (مادة الإدارة في الـ PAT) والعلامات في مواد القانون في السنوات الثلاثة الأولى (أساسيات القانون – قانون التجارة والشركات – تشريعات الأعمال).
7- تحليل (مادة الرياضيات في الـ PAT) والعلامات في مواد المحاسبة في السنوات الثلاثة الأولى (مبادئ المحاسبة– محاسبة إدارية – إدارة مالية).
8- تحليل (مادة الإدارة في الـ PAT) والعلامات في مواد التسويق في السنوات الثلاثة الأولى (مدخل إلى التسويق – إدارة التسويق).
9- تحليل مواد الإنكليزي في السنوات الثلاثة الأولى (إنكليزي 1 – إنكليزي 2 – إنكليزي 3 – إنكليزي 4) لمعرفة مستويات النجاح.
10- تحليل مواد المعلوماتية في السنوات الثلاثة الأولى (تطبيقات المعلوماتية – أساسيات البرمجة – قواعد البيانات) لمعرفة مستويات النجاح.

11- تحليل مواد الرياضيات في السنوات الثلاثة الأولى (رياضيات – إحصاء واحتمالات – رياضيات تطبيقية في الإدارة – أساليب البحث العلمي – بحوث العمليات – ...) لمعرفة مستويات النجاح.

12- تحليل مواد الاقتصاد والتجارة في السنوات الثلاثة الأولى (مدخل إلى علم الاقتصاد – اقتصاد كلي – اقتصاد إسلامي – تجارة دولية) لمعرفة مستويات النجاح.

13- تحليل مواد الإدارة والموارد في السنوات الثلاثة الأولى (مبادئ ووظائف الإدارة – نظريات الإدارة – تنظيم إداري – إدارة الموارد البشرية – سلوك تنظيمي – مسؤولية اجتماعية – إدارة التفاوض) لمعرفة مستويات النجاح.

14- تحليل مواد القانون في السنوات الثلاثة الأولى (أساسيات القانون – قانون التجارة والشركات – تشريعات الأعمال) لمعرفة مستويات النجاح.

15- تحليل مواد المحاسبة في السنوات الثلاثة الأولى (مبادئ المحاسبة– محاسبة إدارية – إدارة مالية) لمعرفة مستويات النجاح.

16- تحليل مواد التسويق في السنوات الثلاثة الأولى (مدخل إلى التسويق – إدارة التسويق) لمعرفة مستويات النجاح.

2.2- سنأخذ تحليل (مادة المعلوماتية في الـ PAT) والعلامات في مواد المعلوماتية في السنوات الثلاثة الأولى مثالا وبقية أنواع التحليل مذكورة بشكل مفصل في الملحق الموجود في الـ CD المرفق.

- الخطوات:

1.2.2- إدخال البيانات لصفحة Excel:

واصفات البيانات في حالة تحليل علامات الـ PAT + 1 2 3:

J	I	H	G	F	E	D	C	B	A	
info	المعدل	قواعد البيانات	أساسيات البرمجة	تطبيقات المعلوماتية	بنية الحاسوب	معلوماتية PAT	الكنية	الاسم	الرقم	1
yes/no										2

مع التنويه أن عمود info يحوي:Yes ناجح في معدل مواد المعلوماتية في السنوات الثلاثة الأولى، No راسب في معدل مواد المعلوماتية في السنوات الثلاثة الأولى.

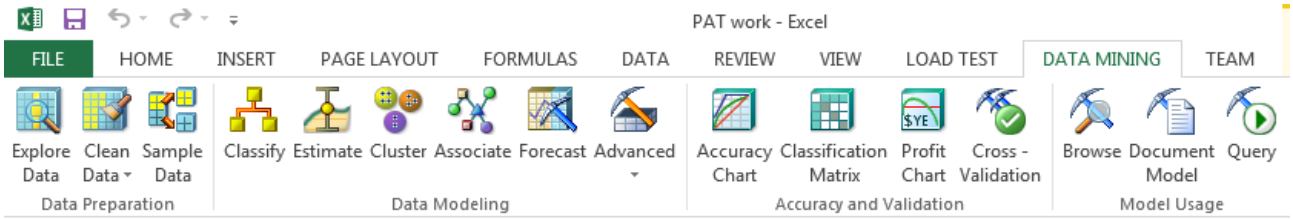
والمعدل هو معدل مواد المعلوماتية (حيث تم ضرب علامة كل مادة من مواد المعلوماتية مع مثلها وجمع العلامات المثقلة، ثم تقسيمها على مجموع أمثال مواد المعلوماتية).

info	المعدل	علامة D ضرب المثل	علامة C ضرب المثل	علامة B ضرب المثل	علامة A ضرب المثل	مادة D	مادة C	مادة B	مادة A
yes	72.5	170	148	146	116	85	74	73	58

2.2.2- تحليل البيانات باستخدام برنامج (Analysis Tool) SQL Server 2008:

أولا- تحديد جدول العلامات الموجود في صفحة الـ Excel.

ثانيا- اختيار خيار Data Mining في شريط الخيارات في Excel:



ثالثا- اختيار نموذج استخراج البيانات (Data Mining Model):

النموذج المستخدم في نظام تحليل العلامات هو Classification التصنيف، لتصنيف علامات الطلاب ضمن قواعد (مثال: علامة المعلوماتية > 11 = yes) وهذه القواعد تفيد في وضع معايير.

رابعا- تحدد الخوارزمية ومتغيرات الخوارزمية من الخيار (Parameters) و المدخلات (input columns) والعمود الذي سيتم عليه التحليل (Column to Analyze):

حيث اخترنا مدخلات الخوارزمية (معلوماتية PAT) والعمود المراد تحليله (تطبيقات المعلوماتية – أساسيات البرمجة – قواعد البيانات).

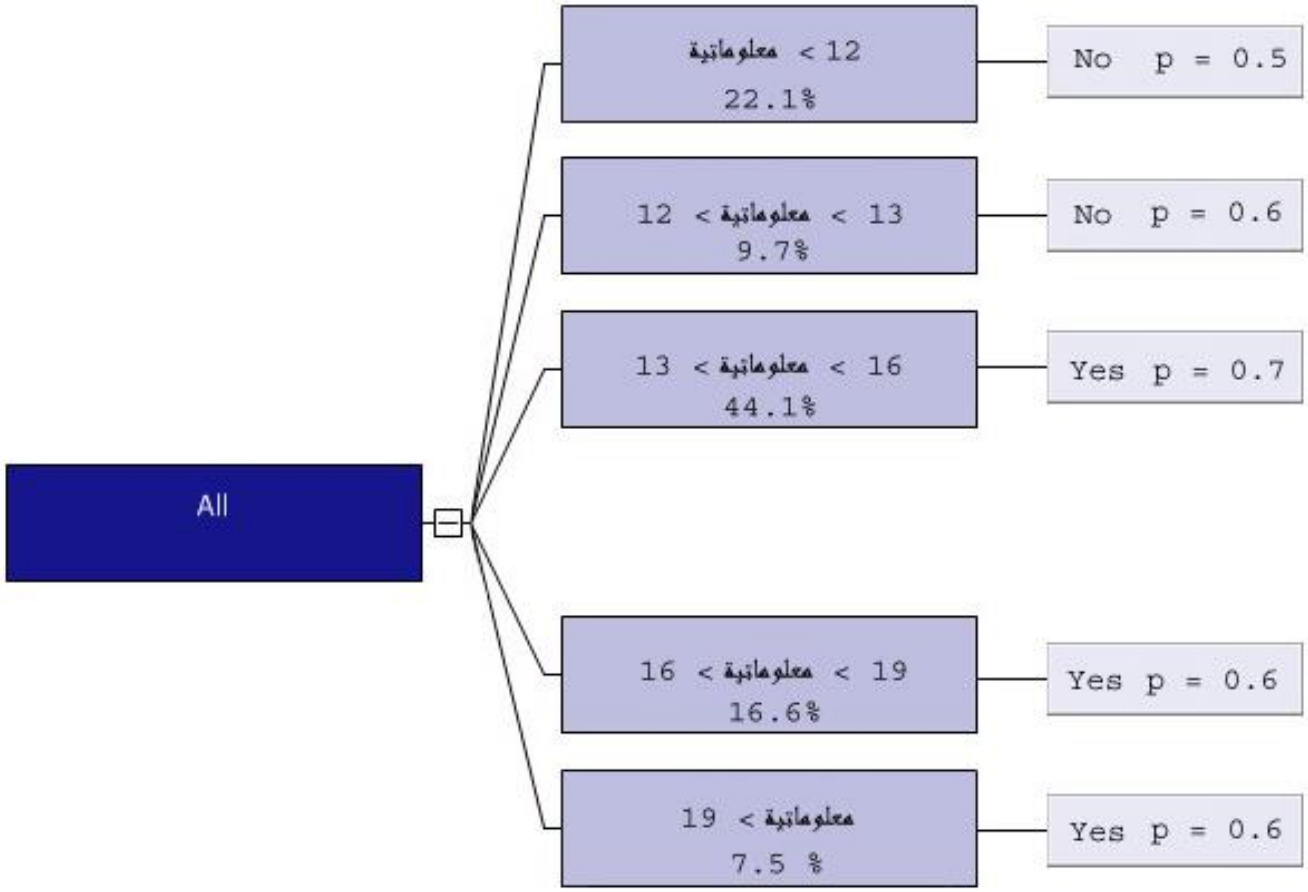
ونختار الـ Minimum_Support = 1 لأننا نريد أخذ كافة العناصر لتحليلها.

و نختار طريقة الحساب الـ Score_Method = 3 طريقة نظرية Bayes.

ونختار نوع الخوارزمية المناسبة: ID3 Decision Tree.

خامسا- بعدها يقوم برنامج SQL Server بإدخال البيانات إلى الخوارزمية وحفظ النتائج في قاعدة بيانات SQL وإظهار نافذة النتائج، القواعد الناتجة عن الخوارزمية وشجرة القرار.

شجرة القرار:



* الـ p هو (Probability) احتمال الـ Yes / No .

من خلال النتائج الناتجة عن الخوارزمية وشجرة القرار نقوم بكتابة القواعد:

(نسمي مادة المعلوماتية x)

No	< 12	x	
No	< 13	x	12 <
Yes	< 16	x	13 <
Yes	< 19	x	16 <
Yes		x	19 <

نستنتج من المجالات السابقة القواعد:

Yes		x	13 <
No	< 13	x	

القواعد:

(علامة المعلوماتية > 13 = Yes) ينجح الطالب في مواد المعلوماتية في السنوات الثلاثة الأولى إذا حصل على 13 فما فوق في مادة المعلوماتية في اختبار القبول PAT.

(علامة المعلوماتية > 13 = Yes) لا ينجح الطالب في مواد المعلوماتية في السنوات الثلاثة الأولى إذا حصل على ما دون الـ 13 في مادة المعلوماتية في اختبار القبول PAT.

3.2.2- استخدام واجهات إدخال Web-Application:

يتم استخدام واجهات إدخال ويب مصممة ببرنامج Visual Studio 2012 لإدخال البيانات الجديدة ومقارنتها مع نموذج الاستخراج (Mining Model) المحفوظ في قاعدة البيانات لإعطاء التنبؤ للبيانات الجديدة.

- الخطوات:

أولا- عند إدخال البيانات الجديدة والضغط على زر (Go) يتم انشاء اتصال ببرنامج SQL Server وانشاء اتصال بقاعدة بيانات SQL.

ثانيا – إنشاء استعلام بالبيانات المدخلة للتنبؤ من نموذج الاستخراج (Mining Model) المنشأ في عملية التحليل.

ثالثا- عرض النتيجة في مربع النص (Yes أو No).

رابعا- عرض احتمال النتيجة في مربع النص (احتمال الـ Yes و الـ No معا).

- الواجهة:

هي واجهة تحليل علامات مواد اختبار الـ PAT وتحليل علامات المواد في السنوات الثلاثة الأولى، حيث يتم إدخال علامة مادة الـ PAT والتنبؤ بالنجاح بالمواد المتعلقة بمادة اختبار الـ PAT في السنوات الثلاثة الأولى.

- التنبؤ باستخدام علامة مادة فحص الـ PAT:
(تحليل علامة مادة الـ PAT وعلامة المواد المتعلقة بها في السنوات الثلاثة الاولى)

علامة الإنكليزي PAT: 16 GO

تحليل مواد الإنكليزي: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.2 Yes With Probability = 0.8

علامة الرياضيات والمعلوماتية PAT: 15 Go

- تحليل مواد الرياضيات: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.2 Yes s With Probability = 0.8

- تحليل مواد المعلوماتية: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.3 Yes With Probability = 0.7

- تحليل مواد المحاسبة: Yes إظهار الشجرة

الاحتمال: NO With Probability = 0.2 Yes With Probability = 0.8

- تحليل مواد الاقتصاد: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.2 Yes With Probability = 0.8

علامة الإدارة PAT: 17 Go

- تحليل مواد الإدارة: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.2 Yes With Probability = 0.8

- تحليل مواد القانون: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.3 Yes With Probability = 0.7

- تحليل مواد التسويق: Yes إظهار الشجرة

الاحتمال: No With Probability = 0.2 Yes With Probability = 0.8

3.2- سنأخذ تحليل مواد المعلوماتية في السنوات الثلاثة الأولى مثالا وبقية أنواع التحليل مذكورة بشكل مفصل في الملحق الموجود في الـ CD المرفق.

- الخطوات:

1.3.2- إدخال البيانات لصفحة Excel:

واصفات البيانات في حالة تحليل علامات الـ PAT + 1 2 3:

J	I	H	G	F	E	D	C	B	A	
info	المعدل	قواعد البيانات	أساسيات البرمجة	تطبيقات المعلوماتية	بنية الحاسوب	معلوماتية PAT	الكنية	الاسم	الرقم	1
yes/no										2

مع التنويه أن عمود info يحوي:Yes ناجح في جميع مواد المعلوماتية في السنوات الثلاثة الأولى، No راسب في جميع مواد المعلوماتية في السنوات الثلاثة الأولى.

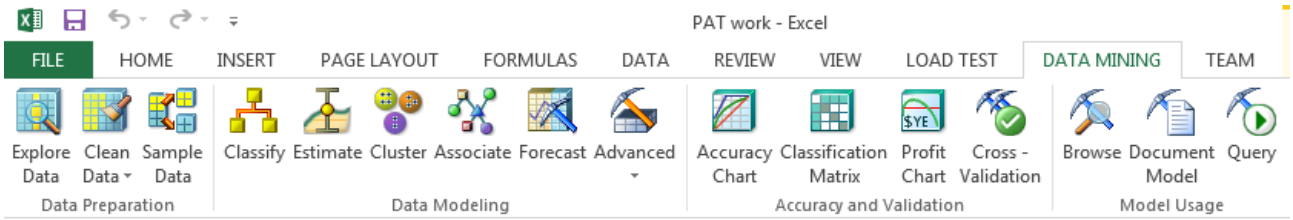
والمعدل هو معدل مواد المعلوماتية (حيث تم ضرب علامة كل مادة من مواد المعلوماتية مع مثلها وجمع العلامات المثقلة، ثم تقسيمها على مجموع أمثال مواد المعلوماتية).

info	المعدل	علامة D ضرب المثل	علامة C ضرب المثل	علامة B ضرب المثل	علامة A ضرب المثل	مادة D	مادة C	مادة B	مادة A
yes	72.5	170	148	146	116	85	74	73	58

2.3.2- تحليل البيانات باستخدام برنامج (Analysis Tool) SQL Server 2008:

أولا- تحديد جدول العلامات الموجود في صفحة الـ Excel.

ثانيا- اختيار خيار Data Mining في شريط الخيارات في Excel:

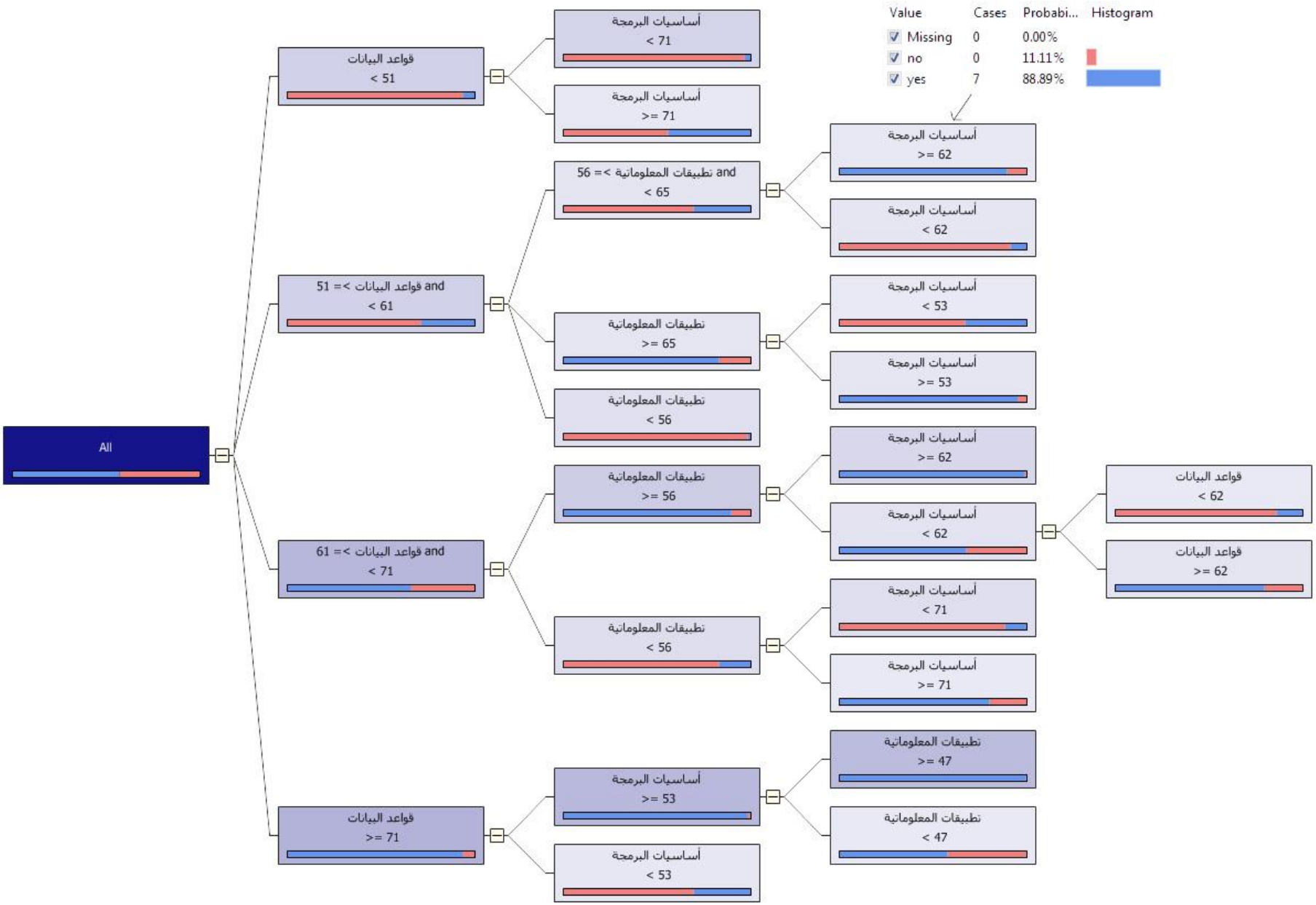


ثالثا- اختيار نموذج استخراج البيانات (Data Mining Model):
النموذج المستخدم في نظام تحليل العلامات هو Classification التصنيف، لتصنيف علامات الطلاب ضمن قواعد (مثال: علامة المعلوماتية > 65 = yes) وهذه القواعد تفيد في وضع معايير.

رابعا- تحدد الخوارزمية ومتغيرات الخوارزمية من الخيار (Parameters) و المدخلات (input columns) والعمود الذي سيتم عليه التحليل (Column to Analyze):
حيث اخترنا مدخلات الخوارزمية (تطبيقات المعلوماتية – أساسيات البرمجة – قواعد البيانات) والعمود المراد تحليله (info).
ونختار الـ Minimum_Support = 1 لأننا نريد أخذ كافة العناصر لتحليلها.
و نختار طريقة الحساب الـ Score_Method = 3 طريقة نظرية Bayes.
ونختار نوع الخوارزمية المناسبة: ID3 Decision Tree.

خامسا- بعدها يقوم برنامج SQL Server بإدخال البيانات إلى الخوارزمية وحفظ النتائج في قاعدة بيانات SQL وإظهار نافذة النتائج، القواعد الناتجة عن الخوارزمية وشجرة القرار.

شجرة القرار:


Value
Cases
Probabi...
Histogram
Missing 0 0.00%
no 0 11.11%
yes 7 88.89%
All
قواعد البيانات
< 51
أساسيات البرمجة
< 71
أساسيات البرمجة
>= 71
51 =< قواعد البيانات and
< 61
تطبيقات المعلوماتية and 56 =<
< 65
أساسيات البرمجة
>= 62
أساسيات البرمجة
< 62
تطبيقات المعلوماتية
>= 65
أساسيات البرمجة
< 53
أساسيات البرمجة
>= 53
تطبيقات المعلوماتية
< 56
61 =< قواعد البيانات and
< 71
تطبيقات المعلوماتية
>= 56
أساسيات البرمجة
>= 62
أساسيات البرمجة
< 62
قواعد البيانات
< 62
قواعد البيانات
>= 62
تطبيقات المعلوماتية
< 56
أساسيات البرمجة
< 71
أساسيات البرمجة
>= 71
قواعد البيانات
>= 71
أساسيات البرمجة
>= 53
تطبيقات المعلوماتية
>= 47
تطبيقات المعلوماتية
< 47
أساسيات البرمجة
< 53

من خلال النتائج الناتجة عن الخوارزمية وشجرة القرار نقوم بكتابة القواعد:

الفرع	المجال الأدنى		المادة	المجال الأعلى	القبول / الرفض
1	1	71 <	قواعد بيانات		?
	2.1		أساسيات البرمجة	53 >	No
	2.2	53 <	أساسيات البرمجة		?
	2.2.1		تطبيقات معلوماتية	47 >	Yes
	2.2.2	47 <	تطبيقات معلوماتية		Yes
2	1	61 <	قواعد بيانات	71 >	?
	2.1		تطبيقات معلوماتية	56 >	?
	2.1.1		أساسيات البرمجة	71 >	No
	2.1.2	71 <	أساسيات البرمجة		Yes
	2.2	56 <	تطبيقات معلوماتية		?
	2.2.1	62 <	أساسيات البرمجة		Yes
	2.2.2		أساسيات البرمجة	62 >	?
	2.2.2.1		قواعد بيانات	62 >	No
	2.2.2.2	62 <	قواعد بيانات		Yes
3	1	51 <	قواعد بيانات	61 >	?
	2.1		تطبيقات معلوماتية	56 >	No
	2.2	56 <	تطبيقات معلوماتية	65 >	?
	2.2.1		أساسيات البرمجة	62 >	No
	2.2.2	62 <	أساسيات البرمجة		Yes
	2.3	65 <	تطبيقات معلوماتية		?
	2.3.1		أساسيات البرمجة	62 >	No
	2.3.2	62 <	أساسيات البرمجة		Yes
4	1		قواعد بيانات	51 >	?
	2.1		أساسيات البرمجة	71 >	No
	2.2	71 <	أساسيات البرمجة		No

نستنتج من المجالات السابقة القواعد:

IF [(61 < قواعد بيانات < 71) and (53 < أساسيات البرمجة)] = No

IF [(71 < قواعد بيانات) and (53 > أساسيات البرمجة) and (47 > تطبيقات معلوماتية)] = Yes

IF [(71 < قواعد بيانات) and (53 > أساسيات البرمجة) and (47 > تطبيقات معلوماتية)] = Yes

IF [(71 < قواعد بيانات) and (56 > تطبيقات معلوماتية) and (71 > أساسيات البرمجة)] = No

IF [(71 < قواعد بيانات) and (56 > تطبيقات معلوماتية) and (71 < أساسيات البرمجة)] = Yes

IF [(71 < قواعد بيانات) and (56 > تطبيقات معلوماتية) and (62 < أساسيات البرمجة)] = Yes

IF [(71 < قواعد بيانات) and (56 > تطبيقات معلوماتية) and (62 > أساسيات البرمجة) and (62 > قواعد بيانات)] = No

IF [(71 < قواعد بيانات) and (56 > تطبيقات معلوماتية) and (62 > أساسيات البرمجة) and (62 < قواعد بيانات)] = Yes

IF [(51 < قواعد بيانات < 61) and (56 > تطبيقات معلوماتية] = No

IF [(51 < قواعد بيانات < 61) and (56 > تطبيقات معلوماتية > 65) and (62 > أساسيات البرمجة)] = No

IF [(51 < قواعد بيانات < 61) and (56 > تطبيقات معلوماتية > 65) and (62 < أساسيات البرمجة)] = Yes

IF [(51 < قواعد بيانات < 61) and (65 > تطبيقات معلوماتية) and (62 > أساسيات البرمجة)] = No

IF [(51 < قواعد بيانات < 61) and (65 > تطبيقات معلوماتية) and (62 < أساسيات البرمجة)] = Yes

IF [(51 > قواعد بيانات) and (71 > أساسيات البرمجة)] = No

IF [(51 > قواعد بيانات) and (71 < أساسيات البرمجة)] = Yes

3.3.2- استخدام واجهات إدخال Web-Application:

يتم استخدام واجهات إدخال ويب مصممة ببرنامج Visual Studio 2012 لإدخال البيانات الجديدة ومقارنتها مع نموذج الاستخراج (Mining Model) المحفوظ في قاعدة البيانات لإعطاء التنبؤ للبيانات الجديدة.

- الخطوات:

أولا- عند إدخال البيانات الجديدة والضغط على زر (Go) يتم انشاء اتصال ببرنامج SQL Server وانشاء اتصال بقاعدة بيانات SQL.

ثانيا – إنشاء استعلام بالبيانات المدخلة للتنبؤ من نموذج الاستخراج (Mining Model) المنشأ في عملية التحليل.

ثالثا- عرض النتيجة في مربع النص (Yes أو No).

رابعا- عرض احتمال النتيجة في مربع النص (احتمال الـ Yes و الـ No معا).

الواجهة:

واجهة تحليل علامات مواد السنوات الثلاثة الأولى، حيث يتم إدخال علامة مادة والتنبؤ بالنجاح في مقررات السنوات الثلاثة الأولى المتعلقة بالمادة المدخلة، والانتقال إلى المرحلة الدراسية الثانية.

الجمهورية العربية السورية
وزارة التعليم العالي
المعهد العالي لإدارة الأعمال

- مواد المعلوماتية :

- تطبيقات المعلوماتية (سنة أولى): 70 GO yes إظهار الشجرة

الاحتمال: no With Probability = 0.132 yes With Probability = 0.868

- أساسيات البرمجة (سنة ثانية): 70 GO yes إظهار الشجرة

الاحتمال: no With Probability = 0.339 yes With Probability = 0.661

- قواعد بيانات (سنة ثالثة): 70 GO yes إظهار الشجرة

الاحتمال: no With Probability = 0.206 yes With Probability = 0.794

- مواد المعلوماتية في السنوات الثلاثة الاولى: GO

yes إظهار الشجرة

احتمال: no With Probability = 0.014 yes With Probability = 0.986

3- **نظام تحليل علامات السنوات 1 2 3 + الاختصاص :**

نقوم بتحليل علامات مواد السنوات الثلاثة الأولى والاختصاص في السنة الرابعة والخامسة، لاستنباط قواعد تفيدنا في وضع معايير لاختيار الاختصاص في السنة الرابعة والخامسة بناء على علامات المواد في السنوات الثلاثة الأولى،

وهذا يفيد في نصح الطالب بالاختصاص المناسب له.

بالإضافة إلى إدخال قيم جديدة والحصول على النتائج (مقارنة البيانات الجديدة مع القواعد الناتجة عن الخوارزمية).

1.3- التحليل:

1- تحليل مادة تطبيقات المعلوماتية في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

2- تحليل مادة أساسيات البرمجة في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

3- تحليل مادة قواعد البيانات في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

4- تحليل مادة رياضيات في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

5- تحليل مادة إحصاء واحتمالات في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

6- تحليل مادة رياضيات تطبيقية في الإدارة في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

7- تحليل مادة أساليب البحث العلمي في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

8- تحليل مادة بحوث العمليات في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

9- تحليل مادة إدارة سلسلة التوريد في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

10- تحليل مادة مدخل إلى علم الاقتصاد في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

11- تحليل مادة اقتصاد كلي في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

12- تحليل مادة اقتصاد إسلامي في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

13- تحليل مادة تجارة دولية في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

14- تحليل مادة مبادئ ووظائف الإدارة في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

15- تحليل مادة نظريات الإدارة في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

16- تحليل مادة تنظيم إداري في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

17- تحليل مادة إدارة الموارد البشرية في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

18- تحليل مادة سلوك تنظيمي في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

19- تحليل مادة مسؤولية اجتماعية في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

20- تحليل مادة إدارة التفاوض في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

21- تحليل مادة مبادئ المحاسبة في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

22- تحليل مادة محاسبة إدارية في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

23- تحليل مادة إدارة مالية في السنة الثالثة واختصاصات الطلاب في السنة الرابعة.

24- تحليل مادة مدخل إلى التسويق في السنة الأولى واختصاصات الطلاب في السنة الرابعة.

25- تحليل مادة إدارة التسويق في السنة الثانية واختصاصات الطلاب في السنة الرابعة.

2.3- سنأخذ تحليل مادة قواعد البيانات مثالا وبقية أنواع التحليل مذكورة بشكل مفصل في الملحق الموجود في الـ CD المرفق.

- الخطوات:

1.2.3- إدخال البيانات لصفحة Excel:

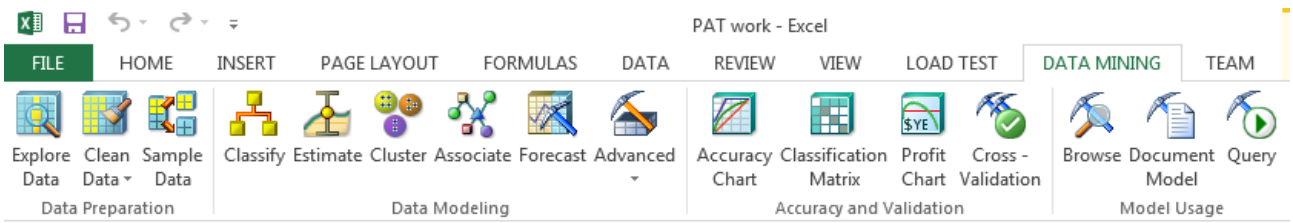
واصفات البيانات في حالة تحليل علامات السنوات 1 2 3 + الاختصاص:

	G	F	E	D	C	B	A
1	الاختصاص	.....	مادة B	مادة A	الكنية	الاسم	الرقم
2	نظم معلومات		80	70	***	***	1
3	موارد بشرية						
4	تسويق						
5	مالية ومصارف						

2.2.3- تحليل البيانات باستخدام برنامج SQL Server 2008 (Analysis Tool):

أولا- تحديد جدول العلامات الموجود في صفحة الـ Excel.

ثانيا- اختيار خيار Data Mining في شريط الخيارات في Excel:



ثالثا- اختيار نموذج استخراج البيانات (Data Mining Model):

النموذج المستخدم في نظام تحليل العلامات هو Classification التصنيف، لتصنيف علامات الطلاب ضمن قواعد (مثال: قواعد بيانات > 90 = نظم معلومات).

رابعا- تحدد الخوارزمية ومتغيرات الخوارزمية من الخيار (Parameters) و المدخلات (input columns) والعمود الذي سيتم عليه التحليل (Column to Analyze):

حيث اخترنا مدخلات الخوارزمية (قواعد البيانات) والعمود المراد تحليله (الاختصاص).

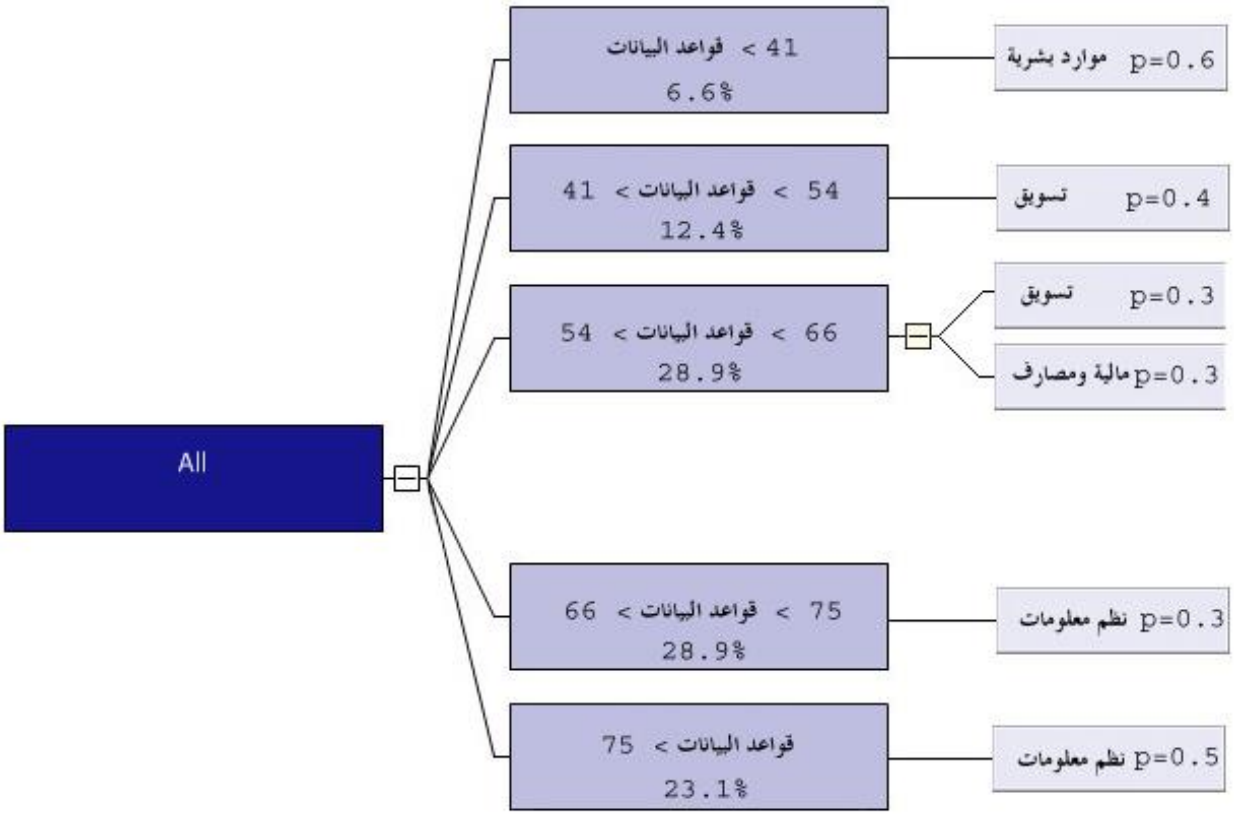
ونختار الـ Minimum_Support = 1 لأننا نريد أخذ كافة العناصر لتحليلها.

و نختار طريقة الحساب الـ Score_Method = 3 طريقة نظرية Bayes.

ونختار نوع الخوارزمية المناسبة: ID3 Decision Tree.

خامسا- بعدها يقوم برنامج SQL Server بإدخال البيانات إلى الخوارزمية وحفظ النتائج في قاعدة بيانات SQL وإظهار نافذة النتائج، القواعد الناتجة عن الخوارزمية وشجرة القرار.

شجرة القرار:



* الـ p هو (Probability) احتمال الـ Yes / No .

من خلال النتائج الناتجة عن الخوارزمية وشجرة القرار نقوم بكتابة القواعد:

(نسمي مادة قواعد البيانات x)

موارد بشرية	< 41	x	
تسويق	< 54	x	41 <
تسويق / مالية ومصارف	< 66	x	54 <
نظم معلومات	< 75	x	66 <
نظم معلومات		x	75 <

القواعد:

If [(41 > قواعد البيانات)] = موارد بشرية

If [(41 < قواعد البيانات < 54)] = تسويق

If [(54 < قواعد البيانات < 66)] = تسويق

If [(54 < قواعد البيانات)] = مالية ومصارف

If [(66 < قواعد البيانات < 75)] = نظم معلومات

If [(75 < قواعد البيانات)] = نظم معلومات

إذا كانت علامة الطالب في مادة قواعد البيانات تحت الـ 41 يكون الاختصاص موارد بشرية ....

3.2.3- استخدام واجهات إدخال Web-Application:

يتم استخدام واجهات إدخال ويب مصممة ببرنامج Visual Studio 2012 لإدخال البيانات الجديدة ومقارنتها مع نموذج الاستخراج (Mining Model) المحفوظ في قاعدة البيانات لإعطاء التنبؤ للبيانات الجديدة.

* الخطوات:

أولا- عند إدخال البيانات الجديدة والضغط على زر (Go) يتم انشاء اتصال ببرنامج SQL Server وانشاء اتصال بقاعدة بيانات SQL.

ثانيا – إنشاء استعلام بالبيانات المدخلة للتنبؤ من نموذج الاستخراج (Mining Model) المنشأ في عملية التحليل.

ثالثا- عرض النتيجة في مربع النص الاختصاص (نظم معلومات أو موارد بشرية أو تسويق أو مالية ومصارف).

رابعا- عرض احتمال النتيجة في مربع النص (احتمال الاختصاصات معا).

- واجهة البرنامج:

الجمهورية العربية السورية
وزارة التعليم العالي
المعهد العالي لإدارة الأعمال

- التنبؤ بالاختصاص: (نصح الطالب بالاختصاص المناسب له بناء على علاماته في السنوات الثلاثة الاولى)

- مواد المعلوماتية:

تطبيقات المعلوماتية (سنة أولى): 60 — مالية ومصارف — إظهار الشجرة
With Probability = 0.26 نظم معلومات
With Probability = 0.26 تسويق
With Probability = 0.20 موارد بشرية
With Probability = 0.30 مالية ومصارف

أساسيات البرمجة (سنة ثانية): 60 — مالية ومصارف — إظهار الشجرة
With Probability = 0.21 نظم معلومات
With Probability = 0.24 تسويق
With Probability = 0.24 موارد بشرية
With Probability = 0.30 مالية ومصارف

قواعد بيانات (سنة ثالثة): 60 — تسويق — مالية ومصارف — إظهار الشجرة
With Probability = 0.12 نظم معلومات
With Probability = 0.32 تسويق
With Probability = 0.24 موارد بشرية
With Probability = 0.32 مالية ومصارف

- مواد الرياضيات:

رياضيات (سنة أولى): 60 — نظم معلومات — إظهار الشجرة
With Probability = 0.36 نظم معلومات
With Probability = 0.21 تسويق
With Probability = 0.19 موارد بشرية
With Probability = 0.24 مالية ومصارف

محاسبة إدارية (سنة ثانية): 60 — تسويق — إظهار الشجرة
With Probability = 0 نظم معلومات
With Probability = 0.53 تسويق
With Probability = 0.29 موارد بشرية
With Probability = 0.24 مالية ومصارف

إدارة مالية (سنة ثالثة): 60 — موارد بشرية — إظهار الشجرة
With Probability = 0.18 نظم معلومات
With Probability = 0.27 تسويق
With Probability = 0.32 موارد بشرية
With Probability = 0.24 مالية ومصارف

- مواد التسويق:

مدخل إلى التسويق (سنة أولى): 60 — مالية ومصارف — إظهار الشجرة
With Probability = 0.13 نظم معلومات
With Probability = 0.27 تسويق
With Probability = 0.20 موارد بشرية
With Probability = 0.40 مالية ومصارف

إدارة التسويق (سنة ثالثة): 60 — تسويق — - - إظهار الشجرة
With Probability = 0.15 نظم معلومات
With Probability = 0.35 تسويق
With Probability = 0.20 موارد بشرية
With Probability = 0.30 مالية ومصارف

لإظهار الاختصاص المناسب اضغط :
إظهار الاختصاص
تسويق

2- النتائج:

1- استنباط قواعد للقبول في المعهد العالي لإدارة الأعمال، ويتم وضع معايير بناء على القواعد الناتجة عن تحليل علامات مواد اختبار الـ PAT.

2- استنباط قواعد لتقييم الطلاب في المرحلة التحضيرية الأولى من خلال تحليل علامات مواد لسنوات الثلاثة الأولى، ليتم وضع معايير استنادا على هذه القواعد.

3- استنباط قواعد لتقييم مستوى الطلاب الداخلين للمعهد العالي لإدارة الأعمال من خلال تحليل علامات مواد اختبار الـ PAT و تحليل علامات المواد في السنوات الثلاثة الأولى المرتبطة بمواد اختبار الـPAT، ليتم وضع معايير استنادا على هذه القواعد.

4- استنباط قواعد ناتجة عن تحليل علامات مواد السنوات الثلاثة الأولى والاختصاصات، ليتم وضع معايير لاختيار الطالب للاختصاص المناسب له بناء على القواعد.

5- دعم القرارات المتعلقة بالقبول والعلامات والخطط الدرسية، استنادا إلى النتائج وشجرات القرار والقواعد الناتجة عن التحليل.

3- التوصيات:

1- العمل على استخدام التنقيب عن المعطيات (Data Mining) بشكل موسع في الأعمال لأهميتها الكبيرة في الوصول إلى المعارف الموجودة في البيانات.

2- إن بناء أنظمة تحليل ودعم القرار يساعد في دعم واتخاذ القرار المناسب، ويساعد في وضع معايير مناسبة.

3- يمكن استخدام التنقيب عن المعطيات (Data Mining) في عملية التوجيه والنصح للطالب في حال لم يدخل الاختصاص الذي نصح به، وبالتالي يمكن تطوير النظام على أن يتوقع المقررات التي سوف يواجه الطالب صعوبة في دراستها.

4- تطوير النظام ليشمل الطلاب في سنوات الاختصاص، عن طريق تحليل علامات الطالب في السنة الرابعة والخامسة.

5- تطوير النظام ليشمل طلاب الماجستير، عن طريق تحليل علامات الطالب في مواد الماجستير.

6- تطوير النظام ليشمل الطلاب في السنوات الخمسة لمعرفة اختصاص الماجستير المناسب.

7- تطوير النظام ليشمل الطلاب الذين قاموا بعمل دورة تدريبية، عن طريق تحليل علامات الستاج وعلامات المواد المتعلقة به.

8- إضافة متغير هام على عملية تحديد الاختصاص وهو رغبة الطالب في اختيار الاختصاص.

4- المراجع:

1- Jiawei Han and Micheline Kamber ,

Data Mining Concepts and Techniques,

Second Edition,

Morgan Kaufmann Publishers,

Publish Date: 2006

2- Jamie MacLennan – ZhaoHui Tang – Bogdan Crivat,

Data Mining With Microsoft SQL Server 2008,

First Edition,

Wiley Publishing, Publish Date: 2009

3- Jeffrey Whitten and Lonnie Bentley,

System Analysis & Design Methods,

Seventh Edition,

McGraw-hill, Publish Date: 2007

4- Alan Dennis - Barbara Haley Wixom - David Tegarden,

System Analysis & Design - With UML Version 2.0 An Object-Oriented Approach,

Third Edition,

Wiley, Publish Date: 2009

5- الملاحق:

موجودة في الـ CD المرفق (لكثر عدد صفحات الملحق 95 صفحة)

تحوي الملاحق تفصيلا عن:

تحليل المواد،

شجرات القرار،

واجهات النظام،

فيديو عن استخدام النظام،

.....