كارلو فيرسيليس

ذكاء الأعمال

تنقيب البيانات وتحسينها لصناعة القرارات

الجزء الأول

ترجمة

م. أحمد فاضل الموسوي

ذكاء الأعمال

ذكاء الأعمال:
تنقيب البيانات وتحسينها لصناعة القرارات

كارلو فيرسيليس

Politecnico di Milano, Italy.

ترجمة

م. أحمد فاضل الموسوي

قسم تكنولوجيا المعلومات

كلية علوم الحاسوب والرياضيات

جامعة ذي قار

Copyright © 2022

الفهرس

الفصل الأول : ذكاء الأعمال **17**

1.1 قرارات فعالة وفي الوقت المناسب **17**

المثال 1.1 - الاحتفاظ في صناعة الهاتف المحمول. 18

مثال 1.2 - التخطيط اللوجستي. 18

1.2 البيانات والمعلومات والمعرفة **20**

المعرفة: 21

1.3 دور النماذج الرياضية **22**

1.4 معماريات ذكاء الأعمال **23**

Data mining : تنقيب البيانات 26

Optimization.: التحسينات 26

Decisions.: القرارات 26

1.4.1 دورة تحليل ذكاء الأعمال 27

التحليل: 28

الرؤية: 28

القرار: 29

التقييم او المراجعة: 29

1.4.2 العوامل التمكينية في مشاريع ذكاء الأعمال 29

التقنيات: 29

التحليلات: 30

الموارد البشرية: 30

1.4.3 تطوير نظام ذكاء الأعمال 31

التحليلات: 31

التصميم: 32

التخطيط: 33

التنفيذ والسيطرة: 33

1.5 الأخلاقيات وذكاء الأعمال **34**

1.6 الملاحظات والقراءات **36**

الفصل الثاني : أنظمة دعم القرار **37**

Decision support systems **37**

2.1 تعريف النظام **37**

2.2 تمثيل عملية صنع القرار **39**

2.2.1 العقلانية وحل المشكلات 40

2.2.2 عملية صنع القرار 42

2.2.3 أنواع القرارات 45

2.2.4 مناهج عملية صنع القرار 50

2.3 تطور نظم المعلومات **51**

2.4 تعريف نظام دعم القرار **53**

2.5 تطوير نظام دعم القرار **57**

2.6 ملاحظات وقراءات **61**

الفصل الثالث: مستودع البيانات **62**

3.1 تعريف مستودع البيانات Definition of data warehouse **62**

3.1.1 سوق/متجر البيانات Data Marts 67

3.1.2 جودة البيانات Data quality 68

3.2 بنية مستودع البيانات Data warehouse architecture **70**

3.2.1 أدوات ETL 72

3.2.2 البيانات الوصفية Metadata 73

3.3 المكعبات والتحليل متعدد الأبعاد **74**

Cubes and multidimensional analysis **74**

3.3.1 التسلسلات الهرمية للمفاهيم وعمليات OLAP 79

3.3.2 تجسيد مكعبات البيانات Materialization of cubes of data 81

3.4 الملاحظات والقراءات **82**

الفصل الرابع النماذج والطرق الرياضية **83**

النماذج الرياضية لصنع القرار **84**

4.1 هيكل النماذج الرياضية **84**

4.2 تطوير نموذج **86**

تحديد المشكلة Problem identification 87

صياغة النموذج Model formulation 87

تطوير الخوارزميات Development of algorithms 89

التنفيذ والاختبار Implementation and test 89

4.3 فئات النماذج **90**

1. النماذج التنبؤية Predictive models 90

2. التعرف على الأنماط ونماذج تعليم الماكنة Pattern recognition and machine learning models 91

3. نماذج التحسين Optimization models 92

4. نماذج إدارة المشروع Project management models 93

5. نماذج تحليل المخاطر Risk analysis models 94

6. نماذج خط الانتظار Waiting line models 94

4.4 ملاحظات وقراءات **96**

الفصل الخامس **97**

التنقيب عن البيانات Data Mining **97**

5.1 تعريف التنقيب عن البيانات **97**

5.1.1 نماذج وأساليب تنقيب البيانات 100

5.1.2 تنقيب البيانات والإحصاءات الكلاسيكية و OLAP 101

5.1.3 تطبيقات التنقيب عن البيانات 102

5.2 تمثيل بيانات الإدخال **103**

5.3 عملية التنقيب عن البيانات **106**

5.4 منهجيات التحليل **112**

5.5 الملاحظات والقراءات **117**

المصادر References **118**

مقدمة المترجم

يسير المجتمع العلمي والصناعي مسرعا و بخطى واسعة نتيجة الحاجة الماسة إلى نتاج البحث العلمي والأكاديمي. فالتقنيات الحديثة وما يعقبها من الحاجة إلى المعرفة الشاملة والدقيقة لمراحل تطبيقها داخل المؤسسات ، تلعب دورا مهما في تطور تلك المؤسسات في سبيل المحافظة على صدارتها في قبال صراع بقاء قائم ونزيه عمن يعطي أفضل الخدمات وبأعلى الجودات. كل هذا لا يمكن لليد العاملة ان تحققه لولا وجود منظومات عمل متكاملة يتداخل في تكوينها وتأسيسها علوم جمة من علوم الحاسبات ، كتحليل النظم وتصميمها، وسياسات أدارة الأعمال الحديثة وتطبيق نظريات الذكاء الأصطناعي في صناعة القرار وأدارة الجودة والمخاطر وغيره. كل هذه الألوان المختلفة من العلوم تمتازج في لوحة فنية رائعة لتنتج لنا موضوعا دراسيا لأدارة المؤسسات تحت مسمى ذكاء الأعمال.

والكتاب الذي بين يديك عزيزي القاريء ، عزيزتي القارئة ، يملك كما هائلا من التفاصيل التي تشرح بخطة محكمة وثابتة الكيفية التي يمكنك فيها فهم النماذج الحديثة التي تقوم على أساسها المؤسسات في أدارة أعمالها. وعلى كل حال، فالكتاب وبنسخته الحالية يقدم المستوى الأولي (كجزء أول) لفهم ذكاء الأعمال بصيغة عامة. الترجمة الحالية تستهدف الفصول الأولى من الكتاب وذلك لحاجة طلبة الدراسات الأولية في علوم الحاسبات وتقنية المعلومات إلى معرفتها أجمالا.

يسعدني أن أستلم مداخلاتكم او تعليقاتكم عن أي شاردة او واردة في الترجمة أو غيره وذلك على البريد الألكتروني almusawiaf@utq.edu.iq.

أحمد فاضل الموسوي

ذي قار - العراق

2022

المقدمة

منذ تسعينيات القرن الماضي ، تمت الإشارة إلى السياق الاجتماعي والاقتصادي الذي يتم فيه تنفيذ الأنشطة الاقتصادية عمومًا باسم مجتمع المعلومات والمعرفة. أدت التغيرات العميقة التي حدثت في أساليب الإنتاج والعلاقات الاقتصادية إلى زيادة أهمية تبادل السلع غير الملموسة ، والتي تتكون في معظمها من عمليات نقل المعلومات.

يرجع تسارع وتيرة عمليات التحول الحالية إلى عاملين. الأول هو العولمة ، التي تُفهم على أنها الترابط المتزايد باستمرار بين اقتصادات مختلف البلدان ، والذي أدى إلى نمو اقتصاد عالمي واحد يتميز بمستوى عالٍ من التكامل.

والثاني هو تقنيات المعلومات الجديدة ، التي تميزت بالانتشار الهائل للإنترنت والأجهزة اللاسلكية ، والتي أتاحت نقل كميات كبيرة من البيانات بسرعة عالية ، وانتشار استخدام وسائل اتصال متطورة.

في هذا السيناريو سريع التطور ، فإن ثروة فرص التنمية غير مسبوقة. توفر سهولة الوصول إلى المعلومات والمعرفة العديد من المزايا للجهات الفاعلة المختلفة في البيئة الاجتماعية والاقتصادية: الأفراد ، الذين يمكنهم الحصول على الأخبار بسرعة أكبر ، والوصول إلى الخدمات بسهولة أكبر وإجراء المعاملات التجارية والمصرفية عبر الإنترنت ؛ الشركات ، التي يمكنها تطوير منتجات وخدمات مبتكرة يمكنها تلبية احتياجات المستخدمين بشكل أفضل ، وتحقيق مزايا تنافسية من خلال الاستخدام الأكثر فعالية للمعرفة المكتسبة ؛ وأخيراً ، الإدارة العامة ، التي يمكنها تحسين الخدمات المقدمة للمواطنين من خلال استخدام تطبيقات الحكومة الإلكترونية ، مثل المدفوعات عبر الإنترنت للمساهمات الضريبية ، وأدوات الصحة الإلكترونية ، من خلال مراعاة التاريخ الطبي لكل مريض ، وبالتالي تحسين جودة خدمات الرعاية الصحية.

في هذا الإطار من التحول الجذري ، تعكس أساليب الحكم داخل المنظمات المعقدة أيضًا التغييرات التي تحدث في البيئة الاجتماعية والاقتصادية ، ويبدو أنها تتأثر بشكل متزايد بالوصول الفوري إلى المعلومات من أجل تطوير خطط عمل فعالة. سيتم استخدام مصطلح المنظمات المعقدة في جميع أنحاء الكتاب للإشارة بشكل جماعي إلى مجموعة متنوعة من الكيانات

العاملة في السياق الاجتماعي والاقتصادي ، بما في ذلك المؤسسات والوكالات الحكومية والمؤسسات المصرفية والمالية والمنظمات غير الهادفة للربح.

أدى اعتماد تقنيات تخزين البيانات الضخمة منخفضة التكلفة والتوافر الواسع لاتصالات الإنترنت إلى إتاحة كميات كبيرة من البيانات التي تم جمعها وتجميعها من قبل المنظمات المختلفة على مر السنين. يمكن للمؤسسات القادرة على تحويل البيانات إلى معلومات ومعرفة استخدامها لاتخاذ قرارات أسرع وأكثر فاعلية وبالتالي تحقيق ميزة تنافسية. على نفس المنوال ، من جانب الإدارة العامة ، يتيح تحليل المعلومات المتاحة تطوير خدمات أفضل ومبتكرة للمواطنين. هذه أهداف طموحة لا يمكن للتكنولوجيا ، مهما كانت متطورة ، أن تؤديها بمفردها ، دون دعم العقول المختصة ومنهجيات التحليل المتقدمة.

هل من الممكن استخراج ، من الكميات الهائلة من البيانات المتاحة ، المعرفة التي يمكن استخدامها من قبل صانعي القرار لمساعدة وتحسين حوكمة المؤسسات والإدارة العامة؟ يمكن تعريف ذكاء الأعمال على أنه مجموعة من النماذج الرياضية ومنهجيات التحليل التي تستغل بشكل منهجي البيانات المتاحة لاسترداد المعلومات والمعرفة المفيدة في دعم عمليات صنع القرار المعقدة.

على الرغم من المعنى المقيد إلى حد ما لمصطلح الأعمال ، والذي يبدو أنه يحصر الموضوع داخل حدود المؤسسات ، فإن أنظمة ذكاء الأعمال تستهدف الشركات وكذلك الأنواع الأخرى من المؤسسات المعقدة ، كما ذكر أعلاه.

منهجيات ذكاء الأعمال متعددة التخصصات وواسعة النطاق ، وتشمل العديد من مجالات التطبيق. في الواقع ، هي تعنى بتمثيل وتنظيم عملية صنع القرار ، وبالتالي مع مجال نظرية القرار ؛ مع جمع وتخزين البيانات التي تهدف إلى تسهيل عملية صنع القرار ، وبالتالي مع تقنيات تخزين البيانات ؛ مع النماذج الرياضية لتحسين واستخراج البيانات ، وبالتالي مع بحوث العمليات والإحصاء؛ أخيرًا ، مع العديد من مجالات التطبيق ، مثل التسويق واللوجستيات والمحاسبة والرقابة والتمويل والخدمات والإدارة العامة.

يمكننا القول أن أنظمة ذكاء الأعمال تميل إلى تعزيز نهج علمي وعقلاني لإدارة المؤسسات والمنظمات المعقدة. حتى استخدام جدول إلكتروني لتقييم الآثار التي تحدثها التقلبات في معدل

الخصم على الميزانية ، على الرغم من بساطته ، يتطلب من جانب صانعي القرار تمثيلًا عقليًا للتدفقات المالية.

توفر بيئة ذكاء الأعمال المعلومات والمعرفة لصانعي القرار المستمدة من معالجة البيانات ، من خلال تطبيق النماذج والخوارزميات الرياضية. في بعض الحالات ، قد تتكون هذه فقط من حساب المجاميع والنسب المئوية ، بينما تستخدم التحليلات الأكثر تطورًا النماذج المتقدمة للتحسين والتعلم الاستقرائي والتنبؤ.

بشكل عام ، يمثل النموذج تجريدًا انتقائيًا لنظام حقيقي ، مصممًا لتحليل وفهم السلوك التشغيلي للنظام الحقيقي من وجهة نظر مجردة. يشتمل النموذج فقط على عناصر النظام التي تعتبر ذات صلة لغرض التحقيق الذي تم إجراؤه. يجدر الاستشهاد بكلمات أينشتاين حول موضوع تطوير النموذج: `` يجب أن يكون كل شيء بسيطًا قدر الإمكان ، ولكن ليس أبسط. " لقد استفادت التخصصات العلمية الكلاسيكية ، مثل الفيزياء ، دائمًا من النماذج الرياضية للتمثيل المجرد لـ الأنظمة الحقيقية ، في حين أن التخصصات الأخرى ، مثل أبحاث العمليات ، قد تعاملت مع تطبيق الأساليب العلمية والنماذج الرياضية لدراسة الأنظمة الاصطناعية ، مثل المؤسسات والمنظمات المعقدة.

"كتاب الطبيعة العظيم" ، كما كتب جاليليو ، "يمكن قراءته فقط من قبل أولئك الذين يعرفون اللغة التي كُتب بها. وهذه اللغة هي الرياضيات. "هل يمكننا أن نطبق أيضًا على تحليل الأنظمة الاصطناعية هذه الرؤية العميقة من أحد الرجال الذين فتحوا الطريق أمام العلم الحديث؟ نحن نعتقد ذلك. في الوقت الحاضر ، فإن القدرات البديهية لصانعي القرار الذين يديرون المؤسسات أو الإدارة العامة يفوقها تعقيد إدارة المنظمات الحالية. كمثال ، ضع في اعتبارك تصميم حملة تسويقية في أسواق ديناميكية وغير متوقعة ، حيث تتوفر ثروة من المعلومات حول السلوك الشرائي للمستهلكين.

اليوم ، من غير المعقول ترك تطبيق نماذج التعلم الاستدلالي المتقدمة جانبًا لاختيار المستفيدين من الحملة ، من أجل تحسين تخصيص الموارد واسترداد إجراء التسويق.

إن تفسير مصطلح ذكاء الأعمال الذي أوضحناه والذي نعتزم تطويره في هذا الكتاب أوسع وأعمق بكثير مقارنة بالمعنى الضيق الذي تم الإعلان عنه خلال السنوات القليلة الماضية من قبل العديد من بائعي البرامج ومجلات تكنولوجيا المعلومات. وفقًا لهذه الرؤية الأخيرة ، يتم تقليل

منهجيات ذكاء الأعمال إلى أدوات إلكترونية للاستعلام والتصور وإعداد التقارير ، بشكل أساسي لأغراض المحاسبة والرقابة.

بالطبع ، لا أحد يستطيع أن ينكر أن الوصول السريع إلى المعلومات هو أداة لا تقدر بثمن بالنسبة لصانعي القرار. ومع ذلك ، فإن هذه الأدوات موجهة نحو تحليلات ذكاء الأعمال ذات الطبيعة السلبية ، حيث صاغ صانع القرار بالفعل في ذهنه بعض المعايير لاستخراج البيانات. إذا كنا نرغب في أن تكون منهجيات ذكاء الأعمال قادرة على التعبير عن إمكاناتها الإستراتيجية الضخمة ، فيجب أن ننتقل إلى أشكال نشطة من الدعم لاتخاذ القرار ، بناءً على التبني المنهجي للنماذج الرياضية القادرة على تحويل البيانات ليس فقط إلى معلومات ولكن أيضًا إلى المعرفة ، و ثم المعرفة إلى ميزة تنافسية فعلية. سيتم البحث في التمييز بين الأشكال السلبية والنشطة للتحليل في الفصل الأول.

قد يعترض المرء على أن الأدوات البسيطة فقط المبنية على مفاهيم فورية وبديهية لها القدرة على إثبات فائدتها في الممارسة. رداً على هذا الاعتراض ، لا يمكننا أن نفعل أفضل من الاقتباس من فلاديمير فابنيك ، الذي ساهم أكثر من أي شخص آخر في تطوير نماذج التعلم الاستقرائي: "لا شيء عملي أكثر من نظرية جيدة". حاولنا خلال هذا الكتاب أن نجعل مراجع متكررة إلى المشكلات والأمثلة المستمدة من التطبيقات الحقيقية من أجل مساعدة القراء على فهم الموضوعات التي تمت مناقشتها ، مع ضمان مستوى مناسب من الدقة المنهجية في وصف النماذج الرياضية.

يصف الجزء الأول المكونات الأساسية التي تشكل بيئة ذكاء الأعمال ، ويناقش هيكل عملية صنع القرار ومراجعة البنى التحتية للمعلومات الأساسية. على وجه الخصوص ، يلخص الفصل الأول إطارًا عامًا لذكاء الأعمال ، ويسلط الضوء على الروابط مع التخصصات الأخرى. يصف الفصل 2 هيكل عملية صنع القرار ويقدم مفهوم نظام دعم القرار ، موضحًا المزايا الرئيسية التي ينطوي عليها ، وعوامل النجاح الحاسمة وبعض قضايا التنفيذ. يعرض الفصل 3 مستودعات البيانات وخزانات البيانات ، أولاً بتحليل الأسباب التي أدت إلى تقديمها ، ثم وصف تحليلات المعالجة التحليلية عبر الإنترنت بناءً على مكعبات متعددة الأبعاد.

الجزء الثاني أكثر منهجية في طبيعته ، ويقدم نظرة عامة شاملة للنماذج الرياضية للتعرف على الأنماط واستخراج البيانات.

يصف الفصل 4 الخصائص الرئيسية للنماذج الرياضية المستخدمة لتحليلات ذكاء الأعمال ، ويقدم تصنيفًا موجزًا للفئات الرئيسية للنماذج. يقدم الفصل الخامس التنقيب عن البيانات ، ويناقش مراحل عملية التنقيب في البيانات وأهدافها. يصف الفصل السادس أنشطة إعداد البيانات لذكاء الأعمال واستخراج البيانات ؛ وتشمل هذه التحقق من صحة البيانات ، واكتشاف الشذوذ ، وتحويل البيانات والحد منها. يقدم الفصل السابع مناقشة تفصيلية لتحليل البيانات الاستكشافية ، التي يتم إجراؤها بواسطة طرق بيانية وإحصاءات موجزة ، من أجل فهم خصائص السمات في مجموعة البيانات وتحديد شدة العلاقات فيما بينها. يصف الفصل الثامن نماذج الانحدار البسيطة والمتعددة ، ويناقش التشخيصات الرئيسية لتقييم أهميتها ودقتها. يوضح الفصل التاسع نماذج لتحليل السلاسل الزمنية ، ودراسة طرق التحلل ، والتمهيد الأسي ونماذج الانحدار الذاتي. الفصل العاشر مخصص بالكامل لنماذج التصنيف ، والتي تلعب دورًا بارزًا في التعرف على الأنماط ونظرية التعلم. بعد وصف معايير التقييم ، يتم توضيح طرق التصنيف الرئيسية ؛ وتشمل هذه الأشجار التصنيف ، وطرق بايز ، والشبكات العصبية ، والانحدار اللوجستي وآلات ناقلات الدعم. يصف الفصل الحادي عشر قواعد الارتباط وخوارزمية Apriori.

يقدم الفصل الثاني عشر أشهر نماذج التجميع: طرق التقسيم ، مثل K-mean و K-medoids ، والطرق الهرمية ، التجميعية والتقسيمية.

يوضح الجزء الثالث تطبيقات استخراج البيانات للتسويق العلائقي (الفصل 13) ، ونماذج تخطيط فريق المبيعات (الفصل الثالث عشر) ، ونماذج تحسين سلسلة التوريد (الفصل الرابع عشر) والطرق التحليلية لتقييم الأداء (الفصل الخامس عشر).

يوفر الملحق "أ" معلومات وروابط لأدوات البرامج المستخدمة لإجراء عمليات التنقيب عن البيانات وتحليلات ذكاء الأعمال الموضحة في الكتاب.

تم إعطاء الأفضلية للبرامج مفتوحة المصدر ، لأنه بهذه الطريقة يمكن للقراء تنزيلها بحرية من الإنترنت للتدرب على الأمثلة المقدمة.

على نفس المنوال ، فإن مجموعات البيانات المستخدمة لتمثيل الموضوعات المختلفة مأخوذة في الغالب من مستودعات في المجال العام. يتضمن الملحق ب وصفًا موجزًا لمجموعات البيانات

المستخدمة في الفصول المختلفة والروابط إلى المواقع التي تحتوي على هذه بالإضافة إلى مجموعات البيانات الأخرى المفيدة للتجربة مع منهجيات التحليل ومقارنتها.

الملاحظات الببليوغرافية في نهاية كل فصل ، بقدر ما هي انتقائية للغاية ، تسلط الضوء على النصوص الأخرى التي وجدناها مفيدة وذات صلة ، بالإضافة إلى المساهمات البحثية ذات القيمة التاريخية المعترف بها.

يستهدف هذا الكتاب ثلاث مجموعات رئيسية من القراء. الأول هم الطلاب الذين يدرسون للحصول على درجة الماجستير في الاقتصاد أو إدارة الأعمال أو غيرها من التخصصات العلمية ، ويحضرون دورة جامعية حول منهجيات ذكاء الأعمال وأنظمة دعم القرار والنماذج الرياضية لصنع القرار. والثاني طلاب في برامج الدكتوراه في تخصصات ذات طبيعة اقتصادية وإدارية. أخيرًا ، قد يكون الكتاب مفيدًا أيضًا للمهنيين الذين يرغبون في تحديث معارفهم والاستفادة من كتاب مرجعي منهجي وعملي. قد يكون القراء الذين ينتمون إلى هذه المجموعة الأخيرة مهتمين بإلقاء نظرة عامة على الفرص التي توفرها أنظمة ذكاء الأعمال ، أو في موضوعات منهجية وتطبيقية محددة يتم تناولها في الكتاب ، مثل تقنيات التنقيب عن البيانات المطبقة على التسويق العلائقي ، ونماذج تخطيط فريق المبيعات ، وسلسلة التوريد نماذج التحسين والطرق التحليلية لتقييم الأداء.

في Politecnico di Milano ، يقود المؤلف مجموعة البحث MOLD - النمذجة الرياضية ، والتحسين ، والتعلم من البيانات ، والتي تجري أنشطة بحثية منهجية على نماذج التعلم الاستقرائي ، والتنبؤ ، والتصنيف ، والتحسين ، وبيولوجيا الأنظمة ، وتحليل الشبكات الاجتماعية ، وكذلك التطبيقية مشاريع حول ذكاء الأعمال والتسويق العلاقي والخدمات اللوجستية. يشتمل موقع الويب الخاص بمجموعة البحث ، www.mold.polimi.it ، على معلومات وأخبار ودراسات متعمقة وروابط مفيدة وتحديثات.

أن كون كتاب ما خالي من الأخطاء المطبعية أمر نادر الحدوث ، خاصة في الطبعة الأولى ، على الرغم من الجهود المبذولة لتلافيها. لذلك ، تم إنشاء منطقة مخصصة للأخطاء والتصويبات على www.mold.polimi.it ، والقراء مدعوون للمساهمة فيها عن طريق إرسال ملاحظة حول أي أخطاء مطبعية قد يجدونها في النص إلى المؤلف على carlo.vercellis @ polimi.it.

أود أن أعرب عن شكري الخاص ل Carlotta Orsenigo ، التي ساعدت في كتابة الفصل العاشر حول نماذج التصنيف وناقشت معي محتوى وتنظيم الفصول المتبقية في الكتاب. كانت مساعدتها في سد الثغرات وتوضيح المفاهيم وتقديم اقتراحات لتحسين النص والأرقام لا تقدر بثمن.

لكتابة هذا الكتاب ، استفدت من تجربتي كمدرس لدورات الدراسات العليا والدراسات العليا. لذلك أود أن أشكر هنا جميع الطلاب العديدين الذين حثوني من خلال أسئلتهم وفضولهم على البحث عن حجج أكثر إقناعًا وثباتًا.

تنشأ العديد من الأمثلة والمراجع عن مشاكل حقيقية من المشاريع التطبيقية التي قمت بها مع مؤسسات ووكالات الإدارة العامة. إنني مدين للعديد من المحترفين لبعض المفاهيم التي أدرجتها في الكتاب: فهي كثيرة جدًا بحيث لا يمكن ذكرها ، لكنها بالتأكيد ستتعرف على نفسها في بعض البيانات ، وأقدم لهم جميعًا شكراً جزيلاً.

جميع الأخطاء المطبعية وعدم الدقة في هذا الكتاب هي مسؤوليتي الخاصة تمامًا.

الجزء الأول
مكونات عملية صنع القرار

الفصل الأول : ذكاء الأعمال

أدى ظهور تقنيات تخزين البيانات منخفضة التكلفة والتوافر الواسع لاتصالات الإنترنت إلى تسهيل وصول الأفراد والمؤسسات إلى كميات كبيرة من البيانات. غالبًا ما تكون مثل هذه البيانات غير متجانسة في الأصل والمحتوى والتمثيل، لأنها تشمل المعاملات التجارية والمالية والإدارية ومسارات تصفح الويب ورسائل البريد الإلكتروني والنصوص والنصوص التشعبية ونتائج الاختبارات السريرية ، على سبيل المثال لا الحصر. أن إمكانية الوصول إليها تفتح سيناريوهات وفرصًا واعدة ، وتثير سؤالًا جذابًا: هل من الممكن تحويل هذه البيانات إلى معلومات ومعرفة يمكن استخدامها من قبل صانعي القرار لمساعدة وتحسين حوكمة الشركات والإدارة العامة؟

يمكن تعريف ذكاء الأعمال على أنه مجموعة من النماذج الرياضية ومنهجيات التحليل التي تستغل البيانات المتاحة لتوليد المعلومات والمعرفة المفيدة لعمليات صنع القرار المعقدة. سيصف هذا الفصل الافتتاحي بشكل عام المشاكل التي ينطوي عليها ذكاء الأعمال ، ويسلط الضوء على الترابط مع التخصصات الأخرى ويحدد المكونات الأساسية النموذجية لبيئة ذكاء الأعمال.

1.1 قرارات فعالة وفي الوقت المناسب

في المنظمات المعقدة ، العامة أو الخاصة ، يتم اتخاذ القرارات على أساس مستمر. قد تكون مثل هذه القرارات أكثر أو أقل أهمية ، ولها تأثيرات طويلة المدى وقصيرة المدى وتشمل أشخاصًا وأدوارًا على مستويات هرمية مختلفة. تعد قدرة عمال المعرفة هؤلاء على اتخاذ القرارات ، كأفراد وكمجتمع ، أحد العوامل الأساسية التي تؤثر على الأداء والقوة التنافسية لمنظمة معينة.

يتوصل معظم العاملين في مجال المعرفة إلى قراراتهم في المقام الأول باستخدام منهجيات سهلة وبديهية ، والتي تأخذ في الاعتبار عناصر محددة مثل الخبرة والمعرفة بمجال التطبيق والمعلومات المتاحة.

يؤدي هذا النهج إلى ركود أسلوب اتخاذ القرار غير المناسب للظروف غير المستقرة التي تحددها التغيرات المتكررة والسريعة في البيئة الاقتصادية. في الواقع ، غالبًا ما تكون عمليات صنع القرار داخل مؤسسات اليوم معقدة وديناميكية للغاية بحيث لا يمكن التعامل معها بشكل فعال من

خلال نهج بديهي ، وتتطلب بدلاً من ذلك موقفًا أكثر صرامة يعتمد على المنهجيات التحليلية والنماذج الرياضية. أشار العديد من المؤلفين مؤخرًا إلى أهمية التحليلات وقيمتها الاستراتيجية في تحديد الميزة التنافسية للمؤسسات ، كما هو موضح في المراجع في نهاية هذا الفصل. يوضح المثالان 1.1 و 1.2 عمليتين شديدتا التعقيد لصنع القرار في ظروف سريعة التغير.

المثال 1.1 - الاحتفاظ في صناعة الهاتف المحمول.

تدرك مديرة التسويق في شركة للهاتف المحمول أن عددًا كبيرًا من العملاء قد توقفوا عن خدمتهم ، تاركين شركتها لصالح مزود منافس. كما يمكن أن نتخيل ، فإن الولاء المنخفض للعملاء ، المعروف أيضًا باسم استنزاف العملاء attrition أو الاضطراب churn،، هو عامل حاسم للعديد من الشركات العاملة في صناعات الخدمات. افترض أن مدير التسويق يمكنه الاعتماد على ميزانية كافية لمتابعة حملة الاحتفاظ بالعملاء التي تستهدف 2000 فرد من إجمالي قاعدة العملاء البالغة 2 مليون شخص. ومن ثم ، فإن السؤال الذي يطرح نفسه بشكل طبيعي هو كيف ينبغي لها أن تختار هؤلاء العملاء ليتم الاتصال بهم من أجل تحسين فعالية الحملة. بمعنى آخر ، كيف يمكن تقدير احتمالية توقف كل عميل عن الخدمة من أجل استهداف أفضل مجموعة من العملاء وبالتالي تقليل الاضطراب وزيادة الاحتفاظ بالعملاء؟ من خلال معرفة هذه الاحتمالات ، يمكن اختيار المجموعة المستهدفة على أنها 2000 شخص لديهم أعلى احتمالية حدوث اضطراب بين العملاء ذوي القيمة التجارية العالية. بدون دعم النماذج الرياضية المتقدمة وتقنيات التنقيب في البيانات ، الموصوفة في الفصل الخامس ، سيكون من الصعب استخلاص تقدير موثوق لاحتمالية التغيير وتحديد المستفيدين الأفضل لحملة تسويقية معينة.

مثال 1.2 - التخطيط اللوجستي.

يرغب مدير اللوجستيات في شركة تصنيع بشرية في تطوير خطة إنتاج لوجستية متوسطة الأجل. هذه عملية صنع قرار شديدة التعقيد تتضمن ، من بين خيارات أخرى ، تخصيص الطلب الناشئ من مناطق السوق المختلفة إلى مواقع الإنتاج ، وشراء المواد الخام والأجزاء المشتراة من الموردين ، وتخطيط الإنتاج للمصانع وتوزيع المنتجات النهائية على مناطق السوق. في شركة تصنيع بشرية نموذجية ، قد يستلزم ذلك عشرات المنشآت ، ومئات الموردين ، وآلاف السلع

والمكونات الجاهزة ، على مدى فترة زمنية مدتها عام واحد مقسمة إلى أسابيع. يشير حجم وتعقيد المشكلة إلى أن نماذج التحسين المتقدمة مطلوبة لابتكار أفضل خطة لوجستية. كما سنرى في الفصل 14 ، تسمح نماذج التحسين بمعالجة المشكلات شديدة التعقيد والواسعة النطاق بنجاح ضمن إطار عمل ذكاء الأعمال.

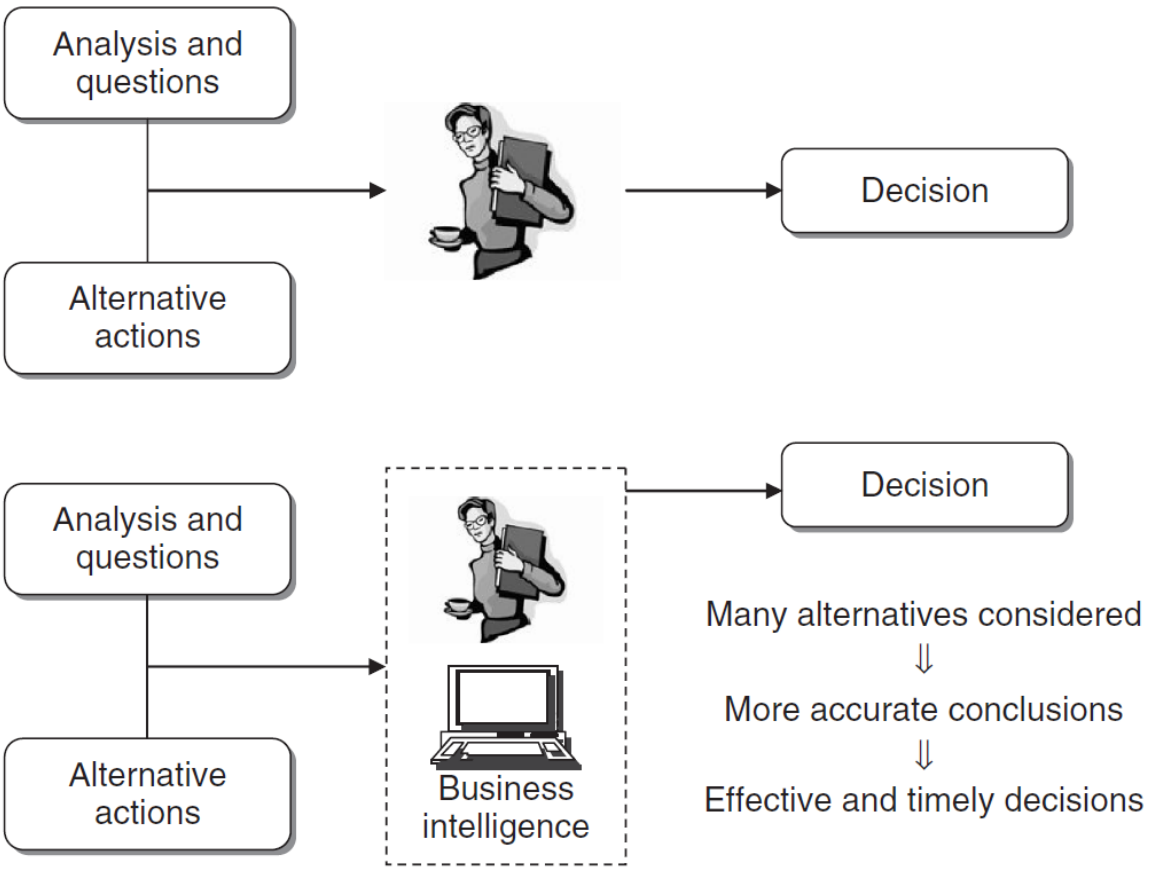
الغرض الرئيسي من أنظمة ذكاء الأعمال هو تزويد العاملين في مجال المعرفة بالأدوات والمنهجيات التي تسمح لهم باتخاذ قرارات فعالة وفي الوقت المناسب.

قرارات فعالة: يسمح تطبيق الأساليب التحليلية الصارمة لصانعي القرار بالاعتماد على المعلومات والمعرفة التي يمكن الاعتماد عليها بشكل أكبر. ونتيجة لذلك ، فهم قادرون على اتخاذ قرارات أفضل ووضع خطط عمل تسمح بتحقيق أهدافهم بطريقة أكثر فعالية. وبالفعل ، فإن اللجوء إلى الأساليب التحليلية الرسمية يجبر صانعي القرار على أن يصفوا بوضوح معايير تقييم الخيارات البديلة والآليات التي تنظم المشكلة قيد البحث. علاوة على ذلك ، يؤدي الفحص المتعمق والفكر اللاحقان إلى إدراك وفهم أعمق للمنطق الكامن وراء عملية صنع القرار.

قرارات في الوقت المناسب: تعمل الشركات في بيئات اقتصادية تتميز بمستويات متزايدة من المنافسة وديناميكية عالية. نتيجة لذلك ، فإن القدرة على الاستجابة بسرعة لأعمال المنافسين وظروف السوق الجديدة هي عامل حاسم في نجاح أو حتى بقاء الشركة.

يوضح الشكل 1.1 الفوائد الرئيسية التي قد تجنيها منظمة معينة من اعتماد نظام ذكاء الأعمال. عند مواجهة مشاكل مثل تلك الموضحة في الأمثلة 1.1 و 1.2 أعلاه ، يسأل صانعو القرار أنفسهم سلسلة من الأسئلة ويطورون التحليل المقابل. ومن ثم ، فإنهم يفحصون ويقارنون عدة خيارات ، ويختارون من بينها القرار الأفضل ، في ظل الظروف المتاحة.

إذا كان بإمكان صانعي القرار الاعتماد على نظام ذكاء الأعمال لتسهيل نشاطهم ، يمكننا أن نتوقع أن الجودة الشاملة لعملية صنع القرار سوف تتحسن بشكل كبير. بمساعدة النماذج والخوارزميات الرياضية ، من الممكن في الواقع تحليل عدد أكبر من الإجراءات البديلة ، وتحقيق استنتاجات أكثر دقة والوصول إلى قرارات فعالة وفي الوقت المناسب. لذلك قد نستنتج أن الميزة الرئيسية المستمدة من اعتماد نظام ذكاء الأعمال توجد في زيادة فعالية عملية صنع القرار.



الشكل 1.1 فوائد نظام ذكاء الأعمال

1.2 البيانات والمعلومات والمعرفة

كما لوحظ أعلاه ، تم تجميع كمية هائلة من البيانات داخل أنظمة المعلومات للمؤسسات العامة والخاصة. تنشأ هذه البيانات جزئيًا من معاملات داخلية ذات طبيعة إدارية ولوجستية وتجارية وجزئيًا من مصادر خارجية. ومع ذلك ، حتى لو تم جمعها وتخزينها بطريقة منهجية ومنظمة ، لا يمكن استخدام هذه البيانات مباشرة لأغراض صنع القرار. يجب معالجتها عن طريق أدوات الاستخراج المناسبة والأساليب التحليلية القادرة على تحويلها إلى معلومات ومعرفة يمكن استخدامها لاحقًا من قبل صانعي القرار.

يمكن فهم الفرق بين البيانات والمعلومات والمعرفة بشكل أفضل من خلال الملاحظات التالية.

البيانات: بشكل عام ، تمثل البيانات تدوينًا منظمًا للكيانات الأولية الفردية ، وكذلك المعاملات التي تنطوي على كيانين أساسيين أو أكثر. على سبيل المثال ، تشير بيانات بائع التجزئة إلى الكيانات الأساسية مثل العملاء ونقاط البيع والأصناف ، بينما تمثل إيصالات المبيعات المعاملات التجارية.

المعلومات : المعلومات هي حصيلة أنشطة الاستخراج والمعالجة التي تتم على البيانات ، وتبدو مفيدة لمن يتلقونها في مجال معين. على سبيل المثال ، بالنسبة لمدير مبيعات شركة بيع بالتجزئة ، فإن نسبة إيصالات المبيعات التي تزيد عن 100 دولار في الأسبوع، أو عدد العملاء الذين يحملون بطاقة ولاء والذين قلوا بنسبة تزيد عن 50٪ من المبلغ الشهري الذي يتم إنفاقه في الأشهر الثلاثة الماضية ، تمثل أجزاء ذات مغزى من المعلومات التي يمكن استخراجها من البيانات الخام المخزنة.

المعرفة:

يتم تحويل المعلومات إلى معرفة عند استخدامها لاتخاذ القرارات وتطوير الإجراءات المقابلة. لذلك ، يمكننا التفكير في المعرفة على أنها تتكون من معلومات يتم وضعها للعمل في مجال معين ، معززة بخبرة وكفاءة صانعي القرار في معالجة وحل المشكلات المعقدة. بالنسبة لشركة البيع بالتجزئة ، قد يكتشف تحليل المبيعات أن مجموعة من العملاء ، الذين يعيشون في منطقة حيث فتح أحد المنافسين مؤخرًا نقطة بيع جديدة ، قد قللوا من حجم أعمالهم المعتاد.

ستؤدي المعرفة المستخرجة بهذه الطريقة في النهاية إلى إجراءات تهدف إلى حل المشكلة المكتشفة ، على سبيل المثال من خلال تقديم خدمة توصيل مجانية جديدة إلى المنازل للعملاء المقيمين في تلك المنطقة المحددة. نود أن نشير إلى أنه يمكن استخلاص المعرفة من البيانات بطريقة سلبية Passive من خلال معايير التحليل التي اقترحها صناع القرار ، أو من خلال التطبيق النشط Active للنماذج الرياضية ، في شكل التعلم الاستقرائي أو التحسين ، كما هو موضح في الفصول التالية.

طورت العديد من المؤسسات والمنظمات العامة والخاصة في السنوات الأخيرة آليات رسمية ومنهجية لجمع وتخزين ومشاركة ثروتها من المعرفة ، والتي يُنظر إليها الآن على أنها أصول غير ملموسة لا تقدر بثمن.

عادة ما يشار إلى نشاط تقديم الدعم للعاملين في مجال المعرفة من خلال تكامل عمليات صنع القرار وتمكين تقنيات المعلومات باسم إدارة المعرفة knowledge management.

من الواضح أن ذكاء الأعمال وإدارة المعرفة يشتركان في درجة من التشابه في أهدافهما. الغرض الرئيسي من كلا التخصصين هو تطوير بيئات يمكن أن تدعم العاملين في مجال المعرفة في عمليات صنع القرار وأنشطة حل المشكلات المعقدة. لرسم حدود بين النهجين ، قد نلاحظ أن منهجيات إدارة المعرفة تركز في المقام الأول على معالجة المعلومات التي عادة ما تكون غير منظمة ، وأحيانًا ضمنية ، ومضمنة في الغالب في المستندات والمحادثات والخبرة السابقة. على العكس من ذلك ، تعتمد أنظمة ذكاء الأعمال على معلومات منظمة ، غالبًا ما تكون ذات طبيعة كمية وعادة ما يتم تنظيمها في قاعدة بيانات. ومع ذلك ، فإن هذا التمييز غامض إلى حد ما: على سبيل المثال ، فإن القدرة على تحليل رسائل البريد الإلكتروني وصفحات الويب من خلال أساليب التنقيب عن النصوص تدفع أنظمة ذكاء الأعمال بشكل تدريجي للتعامل مع المعلومات غير المنظمة.

1.3 دور النماذج الرياضية

يزود نظام ذكاء الأعمال صانعي القرار بالمعلومات والمعرفة المستخرجة من البيانات ، من خلال تطبيق النماذج والخوارزميات الرياضية. في بعض الحالات ، قد يتم تقليل هذا النشاط إلى حسابات المجاميع والنسب المئوية ، ممثلة بيانياً بواسطة مخططات بيانية بسيطة ، بينما تتطلب التحليلات الأكثر تفصيلاً تطوير نماذج التحسين والتعلم المتقدمة.

بشكل عام ، يميل اعتماد نظام ذكاء الأعمال إلى تعزيز نهج علمي وعقلاني لإدارة الشركات والمنظمات المعقدة. حتى استخدام جدول بيانات لتقدير التأثيرات على الميزانية من التقلبات في أسعار الفائدة ، على الرغم من بساطتها ، يجبر صانعي القرار على إنشاء تمثيل عقلي لعملية التدفقات المالية.

لطالما لجأت التخصصات العلمية الكلاسيكية ، مثل الفيزياء ، إلى النماذج الرياضية للتمثيل المجرد للأنظمة الحقيقية. استغلت التخصصات الأخرى ، مثل بحوث العمليات operational research، بدلاً من ذلك تطبيق الأساليب العلمية والنماذج الرياضية لدراسة الأنظمة الاصطناعية ، على سبيل المثال المؤسسات العامة والخاصة. سيصف الجزء الثاني من هذا الكتاب

النماذج الرياضية الرئيسية المستخدمة في معماريات ذكاء الأعمال وأنظمة دعم القرار ، بالإضافة إلى طرق الحلول المقابلة ، بينما يوضح الجزء الثالث العديد من التطبيقات ذات الصلة.

يمكن تلخيص النهج العقلاني النموذجي لتحليل ذكاء الأعمال بشكل تخطيطي في الخصائص الرئيسية التالية.

- أولاً ، يتم تحديد أهداف التحليل ومؤشرات الأداء التي سيتم استخدامها لتقييم الخيارات البديلة.
- يتم بعد ذلك تطوير النماذج الرياضية من خلال استغلال العلاقات بين متغيرات التحكم في النظام والمعلمات ومقاييس التقييم.
- أخيرًا ، يتم إجراء تحليلات ماذا-لو what-if analysis لتقييم التأثيرات على الأداء التي تحددها التغيرات في متغيرات التحكم والتغييرات في المعلمات.

على الرغم من أن هدفهم الأساسي هو تعزيز فعالية عملية صنع القرار ، فإن تبني النماذج الرياضية يوفر أيضًا مزايا أخرى ، والتي يمكن تقديرها بشكل خاص على المدى الطويل. أولاً ، إن تطوير نموذج مجرد يجبر صانعي القرار على التركيز على السمات الرئيسية للمجال الذي تم تحليله ، مما يؤدي إلى فهم أعمق للظاهرة قيد البحث. علاوة على ذلك ، يمكن نقل المعرفة حول المجال المكتسبة عند بناء نموذج رياضي بسهولة أكبر على المدى الطويل إلى أفراد آخرين داخل نفس المنظمة ، مما يسمح بالحفاظ على المعرفة بشكل أكبر مقارنة بعمليات صنع القرار التجريبية. أخيرًا ، النموذج الرياضي الذي تم تطويره لمهمة اتخاذ قرار محددة يكون عامًا ومرنًا لدرجة أنه في معظم الحالات يمكن تطبيقه على مواقف أخرى لاحقة لحل مشاكل من نفس النوع.

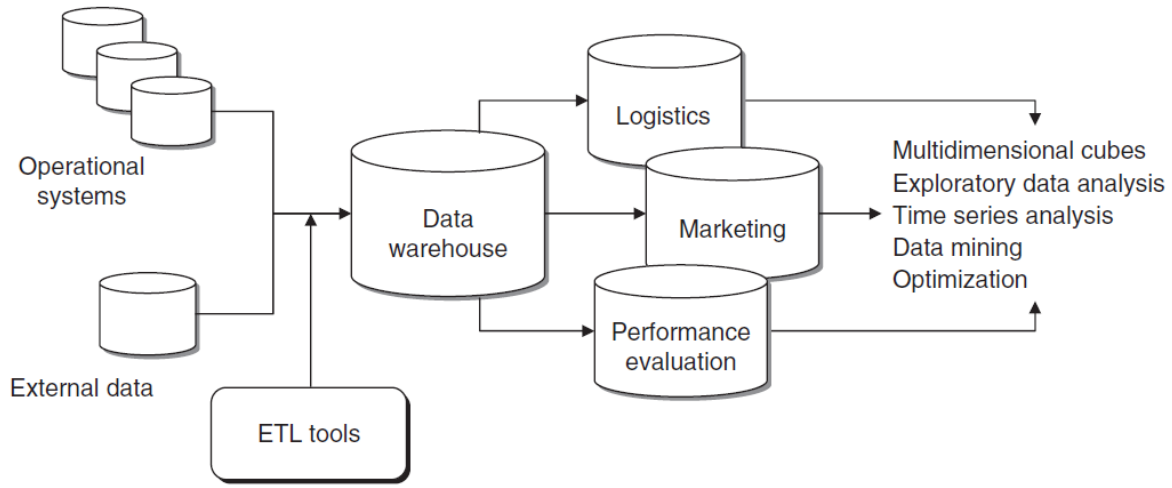
1.4 معماريات ذكاء الأعمال

تشتمل بنية نظام ذكاء الأعمال ، المبينة في الشكل 1.2 ، على ثلاثة مكونات رئيسية.

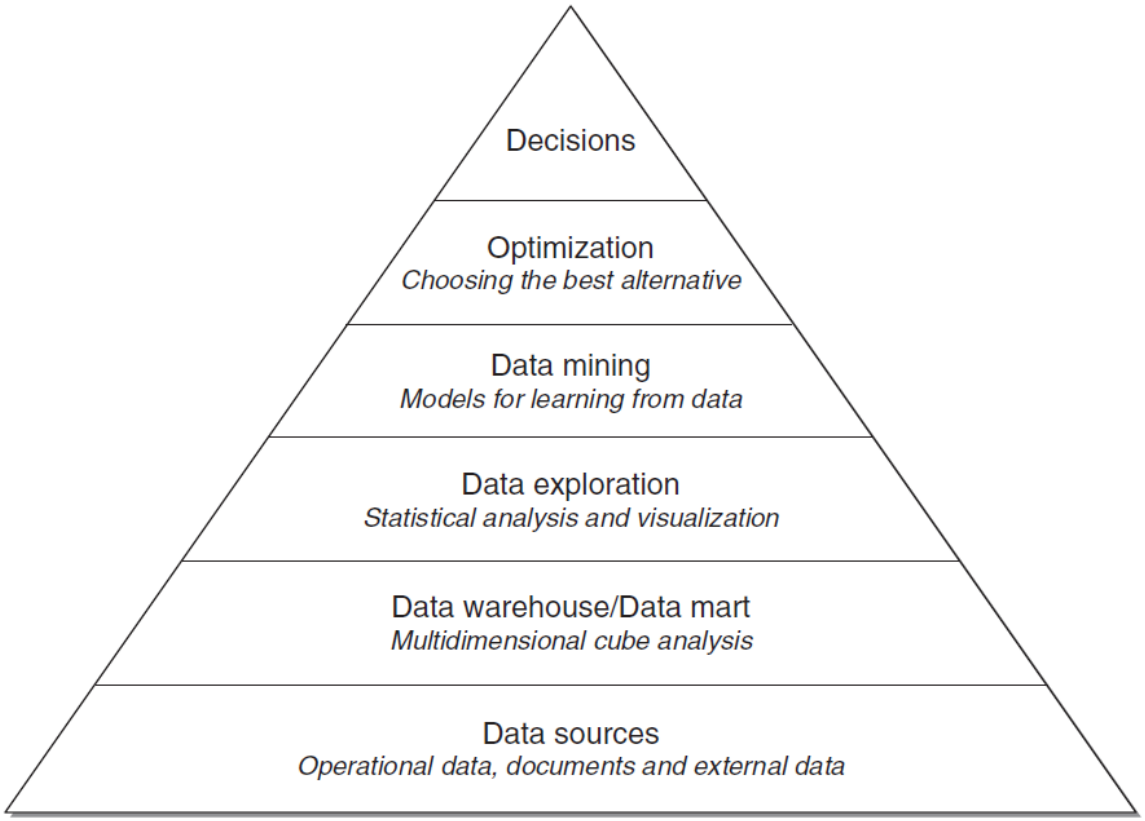
1. **مصادر البيانات.** في المرحلة الأولى ، من الضروري جمع ودمج البيانات المخزنة في مختلف المصادر الأولية والثانوية ، والتي تكون غير متجانسة في الأصل والنوع. تتكون المصادر في معظمها من البيانات التي تنتمي إلى أنظمة التشغيل ، ولكنها قد تتضمن أيضًا مستندات غير منظمة ، مثل رسائل البريد الإلكتروني والبيانات المستلمة من موفرين

خارجيين. بشكل عام ، يلزم بذل جهد كبير لتوحيد ودمج مصادر البيانات المختلفة ، كما هو موضح في الفصل 3.

2. **مستودعات البيانات وسوق البيانات.** باستخدام أدوات الاستخراج والتحويل المعروفة باسم الاستخراج extract, والتحويل transform, والتحميل Load، نموذج (ETL) حيث يتم تخزين البيانات الناشئة من المصادر المختلفة في قواعد بيانات تهدف إلى دعم تحليلات ذكاء الأعمال. يشار إلى قواعد البيانات هذه عادةً باسم مستودعات البيانات data warehouses وسوق البيانات data marts والتي ستكون موضوع الفصل 3.
3. **منهجيات ذكاء الأعمال.** يتم أخيرًا استخراج البيانات واستخدامها لتغذية النماذج الرياضية ومنهجيات التحليل التي تهدف إلى دعم صانعي القرار. في نظام ذكاء الأعمال ، يمكن تنفيذ العديد من تطبيقات دعم القرار ، وسيتم وصف معظمها في الفصول التالية:
   a. تحليل مكعب متعدد الأبعاد multidimensional cube analysis.
   b. تحليل البيانات الاستكشافية exploratory data analysis
   c. تحليل السلاسل الزمنية time series analysis
   d. نماذج التعلم الاستقرائي لاستخراج البيانات inductive learning models for data mining .
   e. نماذج التحسين optimization models.



1.2 بنية نموذجية لذكاء الأعمال الشكل



1.3 المكونات الرئيسية لنظام ذكاء الأعمال

يوضح الهرم في الشكل 1.3 اللبنات الأساسية لنظام ذكاء الأعمال. حتى الآن ، رأينا مكونات المستويين الأولين عند مناقشة الشكل 1.2. ننتقل الآن إلى وصف الطبقات العليا.

استكشاف البيانات Data exploration:

في المستوى الثالث من الهرم نجد الأدوات اللازمة لإجراء تحليل ذكاء الأعمال السلبي ، والذي يتكون من أنظمة الاستعلام وإعداد التقارير ، بالإضافة إلى الأساليب الإحصائية. يشار إلى هذه بالمنهجيات السلبية لأنه يُطلب من صانعي القرار إنشاء فرضيات سابقة أو تحديد معايير استخراج البيانات ، ثم استخدام أدوات التحليل للعثور على إجابات وتأكيد رؤيتهم الأصلية. على سبيل المثال ، ضع في اعتبارك مدير مبيعات الشركة الذي لاحظ أن الإيرادات في منطقة جغرافية معينة قد انخفضت لمجموعة معينة من العملاء. ومن ثم ، قد ترغب في إثبات فرضيتها باستخدام أدوات الاستخراج والتصور ، ثم تطبيق اختبار إحصائي للتحقق من أن استنتاجاتها مدعومة بشكل كافٍ بالبيانات. سيتم وصف الأساليب الإحصائية لتحليل البيانات الاستكشافية في الفصلين 6 و 7.

تنقيب البيانات Data mining :

يتضمن المستوى الرابع منهجيات ذكاء الأعمال النشطة ، والتي تهدف إلى استخراج المعلومات والمعرفة من البيانات.

وتشمل هذه النماذج الرياضية للتعرف على الأنماط ، وتقنيات التعلم الآلي واستخراج البيانات ، والتي سيتم تناولها في الجزء الثاني من هذا الكتاب. على عكس الأدوات التي تم وصفها في المستوى السابق للهرم ، فإن النماذج من النوع النشط لا تتطلب من صانعي القرار صياغة أي فرضية مسبقة ليتم التحقق منها لاحقًا. والغرض منهم بدلاً من ذلك هو توسيع معرفة صانعي القرار.

التحسينات Optimization.:

من خلال الارتقاء بمستوى واحد في الهرم ، نجد نماذج التحسين التي تسمح لنا بتحديد أفضل حل من مجموعة من الإجراءات البديلة ، والتي عادة ما تكون واسعة النطاق إلى حد ما وأحيانًا غير محدودة.

يوضح المثال 1.2 حقلاً نموذجيًا لتطبيق نماذج التحسين. سيتم وصف نماذج التحسين الأخرى المطبقة في التسويق والخدمات اللوجستية في الفصلين 13 و 14.

القرارات Decisions.:

أخيرًا ، يتوافق الجزء العلوي من الهرم مع الاختيار والاعتماد الفعلي لقرار معين ، ويمثل بطريقة ما النتيجة الطبيعية لعملية اتخاذ القرار. حتى عندما تكون منهجيات ذكاء الأعمال متاحة ويتم تبنيها بنجاح ، فإن اختيار القرار يتعلق بصانعي القرار ، الذين قد يستفيدون أيضًا من المعلومات غير الرسمية وغير المهيكلة المتاحة لتكييف وتعديل التوصيات والاستنتاجات التي تم تحقيقها من خلال استخدام النماذج الرياضية.

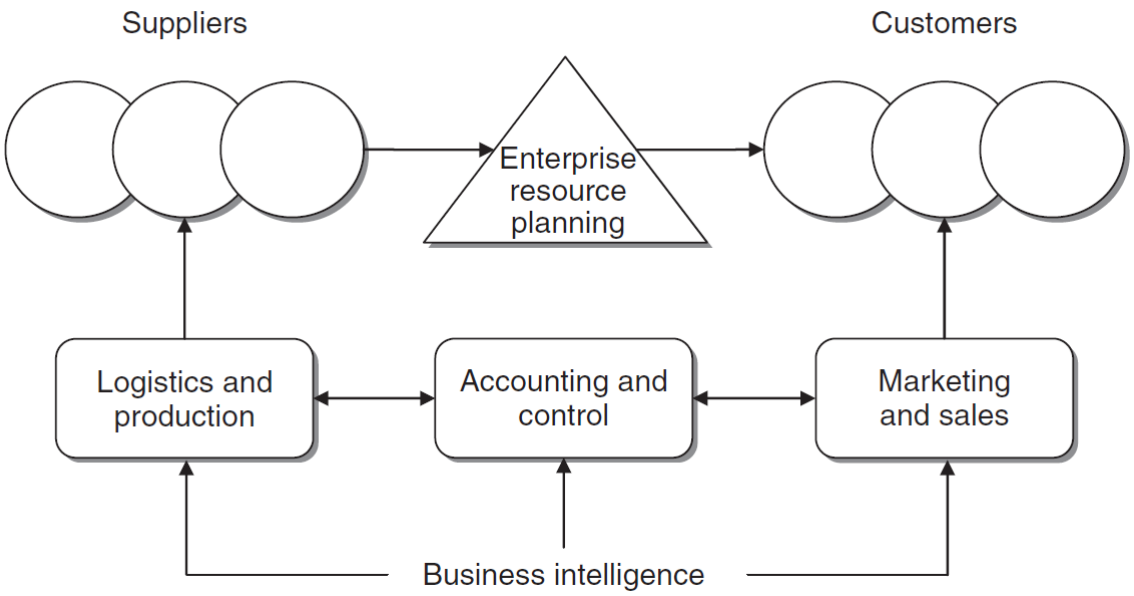
مع تقدمنا من أسفل الهرم إلى قمته ، تقدم أنظمة ذكاء الأعمال بشكل متزايد أدوات دعم أكثر تقدمًا من النوع النشط. حتى الأدوار والكفاءات تتغير. في الجزء السفلي ، يتم توفير الكفاءات المطلوبة في الغالب من قبل متخصصي نظم المعلومات داخل المنظمة ، وعادة ما يشار إليهم بمسؤولي قواعد البيانات. المحللون والخبراء في النماذج الرياضية والإحصائية مسؤولون عن

المراحل الوسيطة. أخيرًا ، تبدو أنشطة صانعي القرار المسؤولين عن مجال التطبيق مهيمنة في الأعلى.

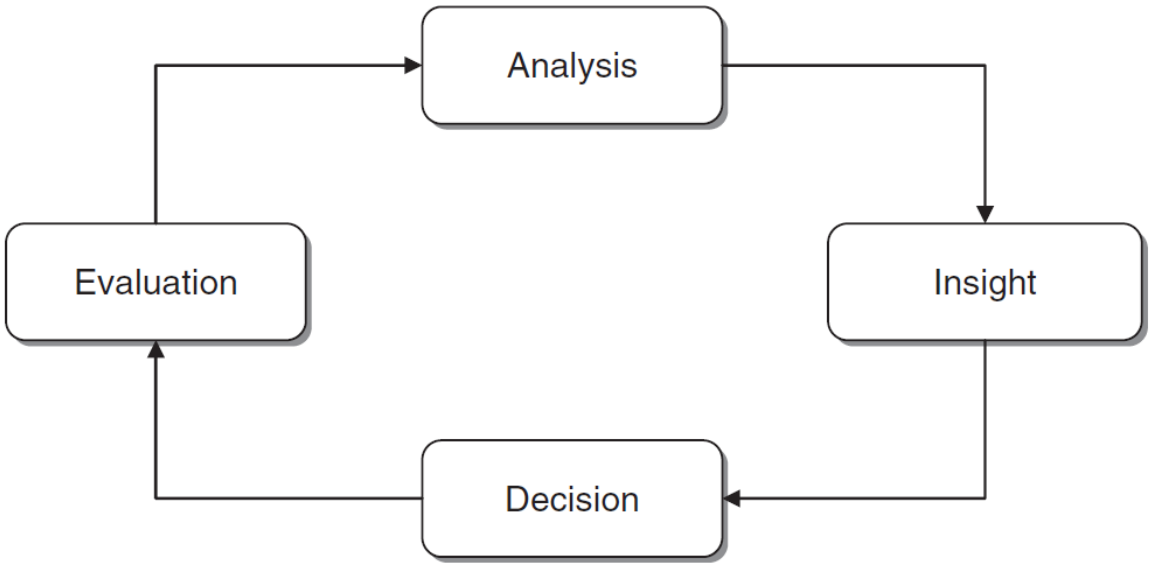
كما هو موضح أعلاه ، تلبي أنظمة ذكاء الأعمال احتياجات أنواع مختلفة من المنظمات المعقدة ، بما في ذلك وكالات الإدارة العامة والجمعيات. ومع ذلك ، إذا قصرنا اهتمامنا على المؤسسات ، فيمكن العثور على منهجيات ذكاء الأعمال بشكل أساسي داخل ثلاثة أقسام من الشركة ، كما هو موضح في الشكل 1.4: التسويق والمبيعات ؛ اللوجستيات والإنتاج ؛ المحاسبة والرقابة. سيتم تخصيص تطبيقات ذكاء الأعمال الموضحة في الجزء الثالث من هذا المجلد بدقة لهذه الموضوعات.

1.4.1 دورة تحليل ذكاء الأعمال

يتبع كل تحليل ذكاء أعمال مساره الخاص وفقًا لمجال التطبيق والموقف الشخصي لصانعي القرار والمنهجيات التحليلية المتاحة. ومع ذلك ، من الممكن تحديد نموذج مسار دوري مثالي والذي يمثل تطور تحليل ذكاء الأعمال النموذجي ، كما هو موضح في الشكل 1.5 ، على الرغم من بقاء الاختلافات الموجودة بناءً على خصوصية كل سياق محدد.



الشكل 1.4 أقسام المؤسسة enterprise المعنية بأنظمة ذكاء الأعمال



الشكل 1.5 دورة مسار تحليل ذكاء الأعمال

التحليل:

أثناء مرحلة التحليل ، من الضروري التعرف على المشكلة المطروحة وتوضيحها بدقة. يجب على صانعي القرار بعد ذلك إنشاء تمثيل عقلي للظاهرة التي يتم تحليلها ، من خلال تحديد العوامل الحاسمة التي يُنظر إليها على أنها الأكثر صلة. قد يساعد توافر منهجيات ذكاء الأعمال بالفعل في هذه المرحلة ، من خلال السماح لصانعي القرار بالتطوير السريع لمسارات التحقيق المختلفة. على سبيل المثال ، يتيح استكشاف مكعبات البيانات في تحليل متعدد الأبعاد ، وفقًا لوجهات نظر منطقية مختلفة كما هو موضح في الفصل 3 ، لصانعي القرار تعديل فرضياتهم بمرونة وسرعة ، حتى يصلوا إلى مخطط تفسير يعتبرونه مرضيًا. وبالتالي ، فإن المرحلة الأولى في دورة ذكاء الأعمال تقود صانعي القرار إلى طرح العديد من الأسئلة والحصول على إجابات سريعة بطريقة تفاعلية.

الرؤية:

تسمح المرحلة الثانية لصانعي القرار بفهم المشكلة المطروحة بشكل أفضل وأعمق ، غالبًا على المستوى السببي. على سبيل المثال ، إذا أظهر التحليل الذي تم إجراؤه في المرحلة الأولى أن عددًا كبيرًا من العملاء يوقفون بوليصة التأمين عند انتهاء الصلاحية السنوية ، فسيكون من الضروري في المرحلة الثانية تحديد الملف الشخصي والخصائص المشتركة بين هؤلاء العملاء. ثم

يتم تحويل المعلومات التي تم الحصول عليها من خلال مرحلة التحليل إلى معرفة خلال مرحلة الرؤية. من ناحية أخرى ، قد يحدث استخراج المعرفة بسبب حدس صانعي القرار ، وبالتالي يعتمد على خبرتهم وربما على معلومات غير منظمة متاحة لهم. من ناحية أخرى ، قد تثبت نماذج التعلم الاستقرائي أيضًا أنها مفيدة جدًا خلال هذه المرحلة من التحليل ، خاصة عند تطبيقها على البيانات المنظمة.

القرار:

خلال المرحلة الثالثة ، يتم تحويل المعرفة التي تم الحصول عليها كنتيجة لمرحلة البصيرة إلى قرارات ومن ثم إلى إجراءات. يتيح توافر منهجيات ذكاء الأعمال تنفيذ مرحلتي التحليل والرؤية بشكل أسرع بحيث يمكن اتخاذ قرارات أكثر فاعلية وفي الوقت المناسب تتناسب بشكل أفضل مع الأولويات الإستراتيجية لمنظمة معينة. يؤدي هذا إلى تقليل إجمالي وقت تنفيذ دورة التحليل - القرار - الإجراء - المراجعة ، وبالتالي إلى عملية صنع القرار ذات الجودة الأفضل.

التقييم او المراجعة:

أخيرًا ، تتضمن المرحلة الرابعة من دورة ذكاء الأعمال قياس الأداء وتقييمه. يجب بعد ذلك وضع مقاييس موسعة لا تقتصر على الجوانب المالية فحسب ، بل تأخذ أيضًا في الاعتبار مؤشرات الأداء الرئيسية المحددة لأقسام الشركة المختلفة. سيصف الفصل 15 منهجيات تحليلية قوية لتقييم الأداء.

1.4.2 العوامل التمكينية في مشاريع ذكاء الأعمال

بعض العوامل أكثر أهمية من غيرها لنجاح مشروع ذكاء الأعمال: التقنيات والتحليلات والموارد البشرية.

1. التقنيات:

تعد تقنيات الأجهزة والبرامج من العوامل التمكينية الهامة التي سهلت تطوير أنظمة ذكاء الأعمال داخل المؤسسات والمنظمات المعقدة. من ناحية أخرى ، زادت قدرات الحوسبة للمعالجات الدقيقة في المتوسط بنسبة 100٪ كل 18 شهرًا خلال العقدين الماضيين ، وانخفضت الأسعار.

لقد أتاح هذا الاتجاه استخدام الخوارزميات المتقدمة المطلوبة لتوظيف أساليب التعلم الاستقرائي ونماذج التحسين ، مع الحفاظ على أوقات المعالجة ضمن نطاق معقول. علاوة على ذلك ، فإنه يسمح باعتماد تقنيات التصور الرسومية الحديثة ، والتي تتميز بالرسوم المتحركة في الوقت الفعلي. هناك عامل تمكين آخر ذي صلة مستمد من الزيادة الهائلة في سعة أجهزة التخزين كبير السعة ، مرة أخرى في انخفاض التكاليف ، مما يمكّن أي مؤسسة من تخزين تيرابايت من البيانات لأنظمة ذكاء الأعمال. ولعب اتصال الشبكة ، في شكل شبكات خارجية أو إنترانت ، دورًا أساسيًا في نشر المعلومات والمعرفة المستخرجة من أنظمة ذكاء الأعمال داخل المنظمات. أخيرًا ، يعد التكامل السهل للأجهزة والبرامج التي تم شراؤها من قبل موردين مختلفين ، أو التي تم تطويرها داخليًا بواسطة منظمة ، عاملاً آخر ذي صلة يؤثر على نشر أدوات تحليل البيانات.

2. التحليلات:

كما هو مذكور أعلاه ، تلعب النماذج الرياضية والمنهجيات التحليلية دورًا رئيسيًا في تعزيز المعلومات واستخراج المعرفة من البيانات المتاحة داخل معظم المنظمات. إن مجرد تصور البيانات وفقًا لوجهات النظر المنطقية المرنة في الوقت المناسب ، كما هو موضح في الفصل 3 ، يلعب دورًا مهمًا في تسهيل عملية صنع القرار ، ولكنه لا يزال يمثل شكلاً من أشكال الدعم السلبي. لذلك ، من الضروري تطبيق نماذج أكثر تقدمًا من التعلم الاستقرائي والتحسين من أجل تحقيق أشكال نشطة من الدعم لعملية صنع القرار.

3. الموارد البشرية:

يتم بناء الأصول البشرية للمؤسسة من خلال كفاءات أولئك الذين يعملون داخل حدودها ، سواء كأفراد أو بشكل جماعي. المعرفة الشاملة التي يمتلكها هؤلاء الأفراد ويتشاركونها تشكل الثقافة التنظيمية. تعد قدرة العاملين في مجال المعرفة على الحصول على المعلومات ثم ترجمتها إلى إجراءات عملية أحد الأصول الرئيسية لأي منظمة ، ولها تأثير كبير على جودة عملية صنع القرار. إذا نفذت مؤسسة معينة نظامًا متقدمًا لذكاء الأعمال ، فلا يزال هناك مجال كبير للتأكيد على المهارات الشخصية للعاملين في مجال المعرفة ، والمطلوبين لإجراء التحليلات وتفسير النتائج ، والعمل على حلول إبداعية وابتكار إجراءات فعالة الخطط. إذا كانت جميع الأدوات

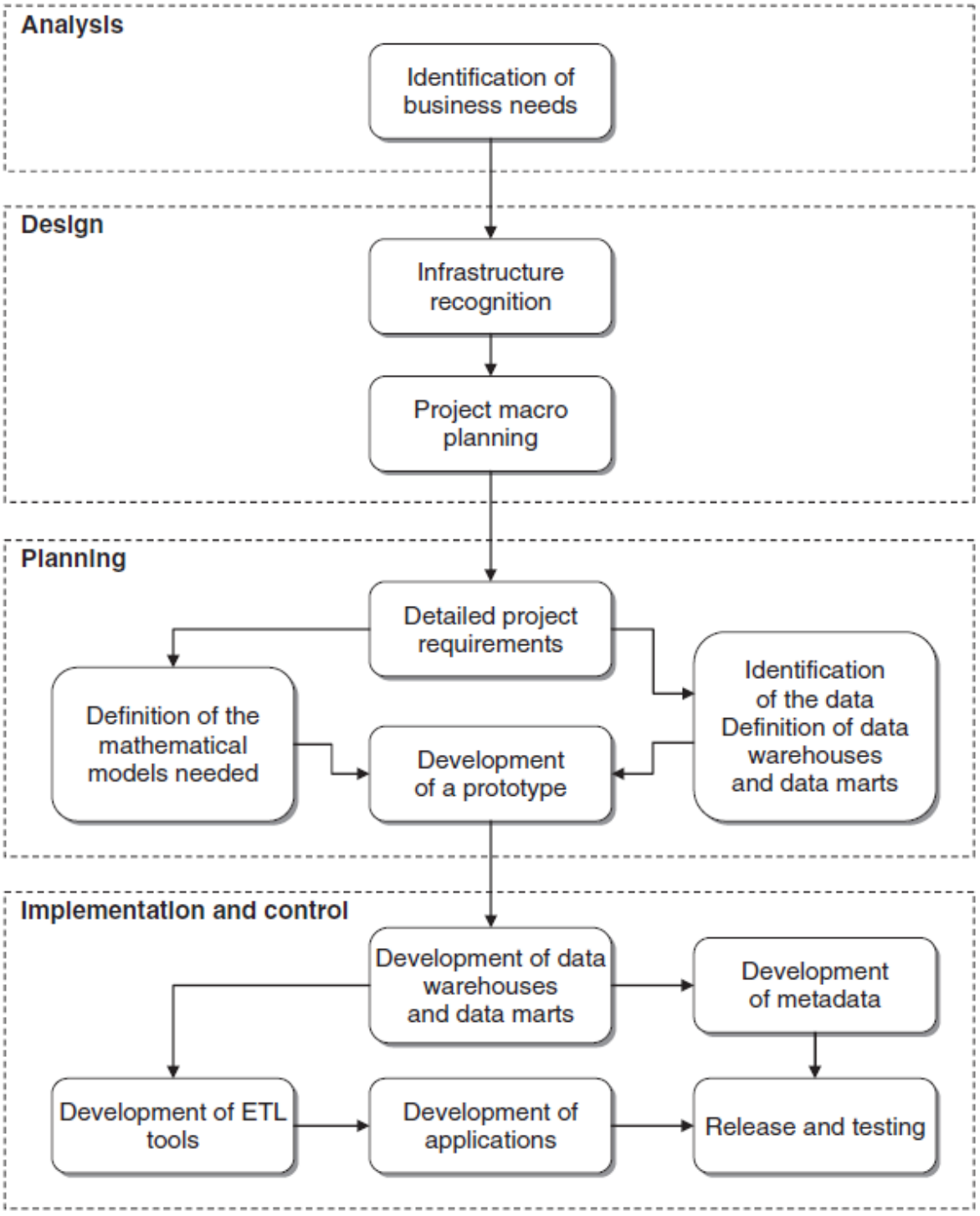
التحليلية المتاحة متساوية ، فإن الشركة التي تستخدم الموارد البشرية التي تتمتع بقدر أكبر من المرونة الذهنية والرغبة في قبول التغييرات في أسلوب صنع القرار ستكون في صالح منافسيها.

1.4.3 تطوير نظام ذكاء الأعمال

يمكن استيعاب تطوير نظام ذكاء الأعمال في مشروع ، بهدف نهائي محدد ، وأوقات تطوير وتكاليف متوقعة ، واستخدام وتنسيق الموارد اللازمة لأداء الأنشطة المخطط لها. يوضح الشكل 1.6 دورة التطوير النموذجية لهندسة ذكاء الأعمال. من الواضح أن المسار المحدد الذي تتبعه كل منظمة قد يختلف عن ذلك الموضح في الشكل. على سبيل المثال ، إذا كانت هياكل المعلومات الأساسية ، بما في ذلك مستودع البيانات وسوق البيانات ، موجودة بالفعل ، فلن تكون المراحل المقابلة المشار إليها في الشكل 1.6 مطلوبة.

التحليلات:

خلال المرحلة الأولى ، يجب تحديد احتياجات المنظمة المتعلقة بتطوير نظام ذكاء الأعمال بعناية. يتم إجراء هذه المرحلة الأولية بشكل عام من خلال سلسلة من المقابلات مع العاملين في مجال المعرفة الذين يؤدون أدوارًا وأنشطة مختلفة داخل المنظمة. من الضروري وصف الأهداف والأولويات العامة للمشروع بوضوح ، وكذلك تحديد التكاليف والفوائد المستمدة من تطوير نظام ذكاء الأعمال.



الشكل 1.6 مراحل تطوير تصميم نظام ذكاء الأعمال

التصميم:

تتضمن المرحلة الثانية مرحلتين فرعيتين وتهدف إلى استنباط مخطط مؤقت للهيكل العام ، مع مراعاة أي تطوير في المستقبل القريب وتطور النظام في المدى المتوسط. أولاً ، من الضروري

إجراء تقييم للبنى التحتية للمعلومات الموجودة. علاوة على ذلك ، يجب فحص عمليات صنع القرار الرئيسية التي يجب أن يدعمها نظام ذكاء الأعمال ، من أجل تحديد متطلبات المعلومات بشكل مناسب. لاحقًا ، باستخدام منهجيات إدارة المشروع الكلاسيكية ، سيتم وضع خطة المشروع ، وتحديد مراحل التطوير ، والأولويات ، وأوقات التنفيذ المتوقعة والتكاليف ، جنبًا إلى جنب مع الأدوار والموارد المطلوبة.

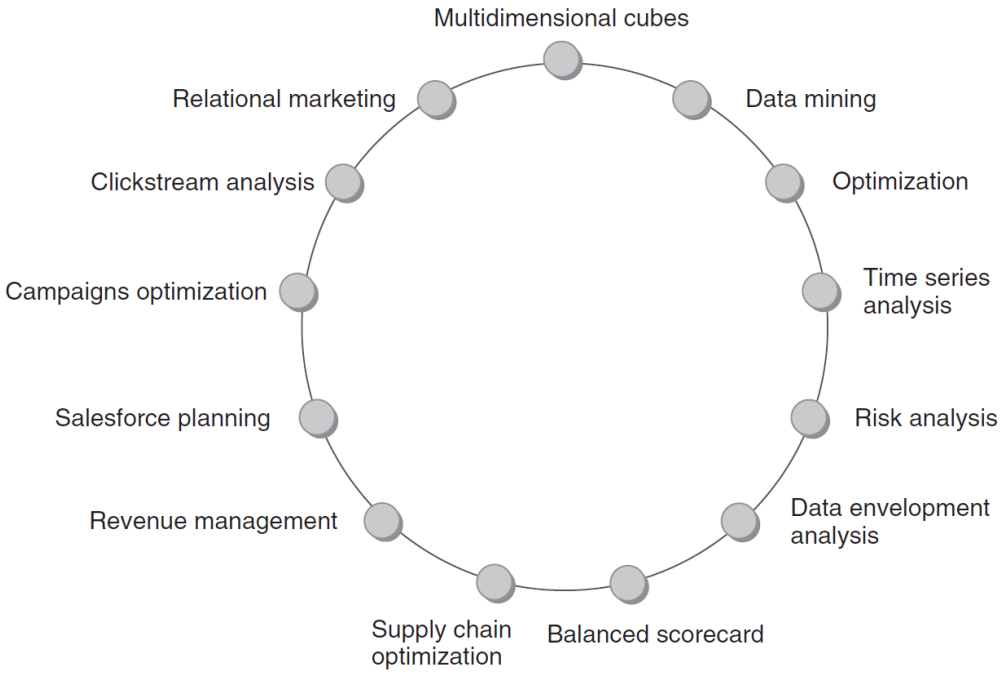
التخطيط:

تتضمن مرحلة التخطيط مرحلة فرعية يتم فيها تحديد وظائف نظام ذكاء الأعمال ووصفها بمزيد من التفصيل. بعد ذلك ، يتم تقييم البيانات الموجودة وكذلك البيانات الأخرى التي يمكن استرجاعها خارجيًا. يسمح هذا بتصميم هياكل المعلومات الخاصة بهندسة ذكاء الأعمال ، والتي تتكون من مستودع بيانات مركزي وربما بعض مجموعات بيانات الأقمار الصناعية. بالتزامن مع التعرف على البيانات المتاحة ، يجب تحديد النماذج الرياضية التي سيتم اعتمادها ، وضمان توافر البيانات المطلوبة لتغذية كل نموذج والتحقق من أن كفاءة الخوارزميات التي سيتم استخدامها ستكون كافية لحجم الناتج الناتج مشاكل. أخيرًا ، من المناسب إنشاء نموذج أولي للنظام ، بتكلفة منخفضة وبقدرات محدودة ، من أجل الكشف مسبقًا عن أي تباين بين الاحتياجات الفعلية ومواصفات المشروع.

التنفيذ والسيطرة:

تتكون المرحلة الأخيرة من خمس مراحل فرعية رئيسية. أولاً ، تم تطوير مستودع البيانات وكل سوق بيانات محدد. تمثل هذه البنى التحتية للمعلومات التي ستغذي نظام ذكاء الأعمال. من أجل شرح معنى البيانات الموجودة في مستودع البيانات والتحويلات المطبقة مسبقًا على البيانات الأولية ، يجب إنشاء أرشيف للبيانات الوصفية ، كما هو موضح في الفصل 3. علاوة على ذلك ، تم إعداد إجراءات ETL لاستخراج وتحويل البيانات الموجودة في المصادر الأولية ، وتحميلها في مستودع البيانات وسوق البيانات. تهدف الخطوة التالية إلى تطوير تطبيقات ذكاء الأعمال الأساسية التي تسمح بإجراء التحليلات المخطط لها. أخيرًا ، تم تحرير النظام للاختبار والاستخدام.

يقدم الشكل 1.7 نظرة عامة على المنهجيات الرئيسية التي يمكن تضمينها في نظام ذكاء الأعمال ، والتي سيتم وصف معظمها في الفصول التالية. بعضها له طبيعة منهجية ويمكن استخدامه عبر مجالات تطبيق مختلفة ، بينما لا يمكن تطبيق البعض الآخر إلا على مهام محددة.



الشكل 1.7 حافظة المنهجيات المتاحة في نظام ذكاء الأعمال

1.5 الأخلاقيات وذكاء الأعمال

يثير اعتماد منهجيات ذكاء الأعمال وطرق استخراج البيانات وأنظمة دعم القرار بعض المشكلات الأخلاقية التي لا ينبغي التغاضي عنها. في الواقع ، فإن التقدم نحو مجتمع المعلومات والمعرفة يفتح فرصًا لا حصر لها ، ولكنه قد يؤدي أيضًا إلى حدوث تشوهات ومخاطر ينبغي تجنبها وتجنبها باستخدام قواعد وآليات رقابة مناسبة. لا ينبغي التسامح مع استخدام البيانات من قبل المؤسسات العامة والخاصة بشكل غير لائق ولا يحترم حق الأفراد في الخصوصية.

بشكل عام ، يجب أن نحذر من النمو المفرط للقوة السياسية والاقتصادية للمؤسسات مما يسمح لعمليات التحول الموضحة أعلاه بالفائدة الحصرية ومن جانب واحد لهذه الشركات نفسها ، على حساب المستهلكين والعاملين وسكان النظام البيئي للأرض.

ومع ذلك ، حتى في حالة فشل اللوائح المحددة التي من شأنها منع إساءة استخدام جمع البيانات والتحقيقات الغازية ، فمن الضروري أن يلتزم محللو استخبارات الأعمال وصناع القرار بالمبدأ الأخلاقي لاحترام الحقوق الشخصية للأفراد. إن خطر تجاوز الحد الفاصل بين الاستخدام الصحيح والمتطفل للمعلومات مرتفع بشكل خاص في مجالات التسويق العلائقي والتعدين على الويب ، كما هو موضح في الفصل 13. على سبيل المثال ، حتى إذا تم إخفاءه تحت أسماء تبدو غير مؤذية مثل "إثراء البيانات" ، فإن المعلومات الخاصة حول يتداول الأفراد والأسر ، لكن هذا لا يعني أنه من الأخلاقي أن يستخدمه صناع القرار والشركات.

إن احترام الحق في الخصوصية ليس هو القضية الأخلاقية الوحيدة المتعلقة باستخدام أنظمة ذكاء الأعمال. كان هناك الكثير من النقاش في السنوات الأخيرة حول المسؤوليات الاجتماعية للشركات ، مما أدى إلى إدخال المفهوم الجديد لأصحاب المصلحة. يشير هذا المصطلح إلى أي شخص لديه أي مصلحة في أنشطة مؤسسة معينة ، مثل المستثمرين والموظفين والنقابات العمالية والمجتمع المدني ككل. هناك تنوع في الآراء حول ما إذا كان يتعين على الشركة السعي لتحقيق أقصى قدر من الأرباح على المدى القصير ، والعمل حصريًا لصالح المساهمين ، أو يجب عليها بدلاً من ذلك اتباع نهج يأخذ في الاعتبار العواقب الاجتماعية لقراراتها.

نظرًا لأن هذا ليس المكان المناسب لمناقشة مشكلة بهذا الحجم ، سنقتصر على الإشارة إلى أن التحليلات المستندة إلى أنظمة ذكاء الأعمال تتأثر بهذه المشكلة ، وبالتالي فهي تتعرض لخطر استخدامها لتحقيق أقصى قدر من الأرباح حتى عندما يجب أن تكون هناك اعتبارات مختلفة تسود فيما يتعلق بالعواقب الاجتماعية للقرارات المتخذة ، وفقًا لمنطق نعتقد أنه يجب رفضه. على سبيل المثال ، هل من الصواب تطوير نموذج تحسين بهدف توزيع التكاليف على نطاق دولي من أجل الالتفاف على أنظمة الضرائب في بعض البلدان؟ هل من المشروع اتخاذ قرار بشأن الوضع الأمثل للخزان في السيارة لتقليل تكاليف الإنتاج ، حتى لو كان ذلك قد يتسبب في ضرر جسيم للركاب في حالة حدوث تصادم؟ كما ثبت من خلال هذه الأمثلة ، فإن المحللين الذين يطورون نموذجًا رياضيًا وأولئك الذين يتخذون القرارات لا يمكن أن يظلوا محايدين ، لكن لديهم التزامًا أخلاقيًا لاتخاذ موقف أخلاقي.

1.6 الملاحظات والقراءات

كما لوحظ أعلاه ، تعد منهجيات ذكاء الأعمال متعددة التخصصات بطبيعتها ولم يبدأ المجتمع العلمي في التعامل معها إلا مؤخرًا كموضوع منفصل. ونتيجة لذلك ، تم إصدار معظم المنشورات في السنوات الأخيرة في شكل تقارير صحفية أو ترويجية ، مع استثناءات قليلة. فيما يلي بعض القراءات المقترحة: Moss and Atre (2003) ، يقدمان وصفًا للمبادئ التوجيهية التي يجب اتباعها في تطوير أنظمة ذكاء الأعمال ؛ Simon and Shaffer (2001) حول تطبيقات ذكاء الأعمال للتجارة الإلكترونية ؛ Kudyba و Hoptroff (2001) لمقدمة عامة للموضوع ؛ وأخيرًا ، جيوفينازو (2002) ومارشال وآخرون. (2004) يركز على تطبيقات ذكاء الأعمال عبر الإنترنت. تمت الإشارة مؤخرًا إلى الدور الاستراتيجي للطرق التحليلية ، في شكل نماذج رياضية تنبؤية وتحسينية ، من قبل عدد من المؤلفين ، من بينهم دافنبورت وهاريس (2007) وأيريس (2007).

تم فحص تكامل معماريات ذكاء الأعمال وأنظمة دعم القرار وإدارة المعرفة بواسطة Bolloju et al. (2002) ، نعمتي وآخرون. (2002) و Malone et al. (2003). حجم راسموسن وآخرون. (2002) يصف دور منهجيات ذكاء الأعمال للتطبيقات المالية ، والتي لم يتم تناولها في هذا النص. للحصول على اعتبارات ذات طبيعة عامة بشأن الآثار الأخلاقية لقرارات الشركات ، انظر باكان (2005). يفحص Snapper (1998) الجوانب الأخلاقية التي ينطوي عليها تطبيق منهجيات ذكاء الأعمال في القطاع الطبي.

الفصل الثاني : أنظمة دعم القرار

Decision support systems

نظام دعم القرار (DSS) هو تطبيق حاسوبي تفاعلي يجمع بين البيانات والنماذج الرياضية لمساعدة صانعي القرار في حل المشكلات المعقدة التي تواجههم في إدارة المؤسسات والمؤسسات العامة والخاصة. كما هو موضح في الفصل 1 ، يمكن اعتبار أدوات التحليل التي توفرها بنية ذكاء الأعمال على أنها **نظام دعم القرار** قادرة على تحويل البيانات إلى معلومات ومعرفة مفيدة لصانعي القرار. في هذا الصدد ، تعد **أنظمة دعم القرار** مكونًا أساسيًا في تطوير بنية ذكاء الأعمال.

في هذا الفصل سوف نناقش أولاً هيكل عملية صنع القرار. علاوة على ذلك ، سيتم عرض تطور نظم المعلومات بإيجاز. سنقوم بعد ذلك بتعريف نظام دعم القرار ، مع تحديد المزايا الرئيسية والإشارة إلى عوامل النجاح الحاسمة المتعلقة بتقديمها. أخيرًا ، سيتم وصف مراحل تطوير مشروع نظام دعم القرار ، ومعالجة القضايا الأكثر صلة فيما يتعلق بتنفيذه.

2.1 تعريف النظام

غالبًا ما يستخدم مصطلح النظام في اللغة اليومية: على سبيل المثال ، نشير إلى النظام الشمسي أو الجهاز العصبي أو نظام العدالة. تشترك الكيانات التي نطلق عليها بشكل حدسي الأنظمة في خاصية مشتركة ، والتي سنعتمدها كتعريف تجريدي لمفهوم النظام: كل منها يتكون من مجموعة من المكونات التي ترتبط ببعضها البعض بطريقة ما لتوفير نتيجة جماعية واحدة وهدف مشترك.

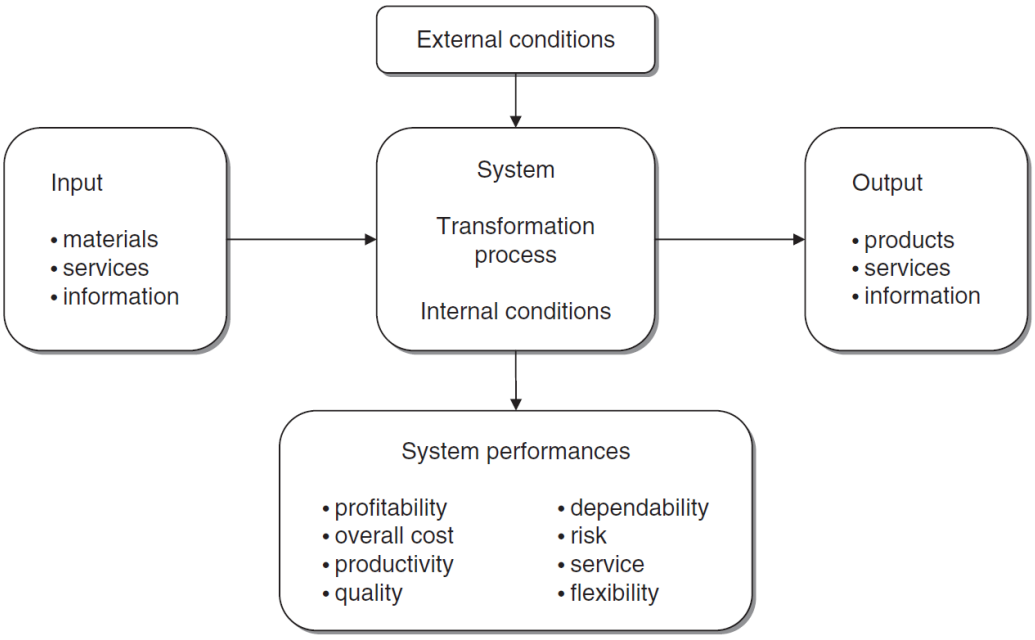
يتميز كل نظام بحدود تفصل مكوناته الداخلية عن البيئة الخارجية. يُقال أن النظام مفتوح إذا كان من الممكن عبور حدوده في كلا الاتجاهين عن طريق تدفقات المواد والمعلومات. في حالة عدم وجود مثل هذه التدفقات ، يقال إن النظام مغلق. بشكل عام ، يتلقى أي نظام تدفقات مدخلات محددة ، وينفذ عملية تحويل داخلي ويولد تدفقات مخرجات يمكن ملاحظتها.

كما يمكن أن تتخيل ، يمكن استخدام هذا التعريف المجرد للنظام لوصف فئة واسعة من ظواهر العالم الحقيقي. على سبيل المثال ، **الهيكل اللوجستي للمؤسسة** هو نظام يتلقى كمدخلات مجموعة من المواد والخدمات والمعلومات ويعود كمخرج لمجموعة من المنتجات والخدمات والمعلومات. بشكل أكثر عمومية ، حتى المؤسسة ، كليا أو جزئيا ، يمكن تمثيلها كنظام ، بشرط أن تكون الحدود وكذلك تدفقات المدخلات والمخرجات محددة بوضوح.

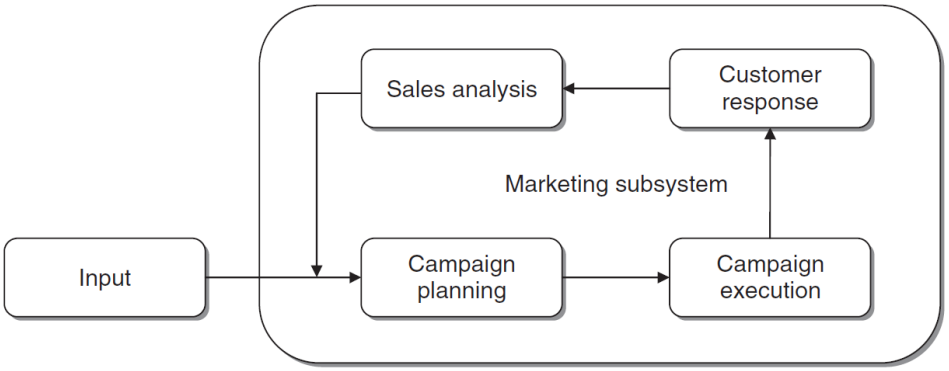
يوضح الشكل 2.1 الهيكل الذي سنستخدمه كمرجع لوصف مفهوم النظام. يتلقى النظام مجموعة من تدفقات المدخلات ويعيد مجموعة من تدفقات المخرجات من خلال عملية تحويل تنظمها الظروف الداخلية والظروف الخارجية. يتم تقييم فعالية وكفاءة النظام باستخدام مؤشرات أداء قابلة للقياس يمكن تصنيفها إلى فئات مختلفة. يوضح الشكل الأنواع الرئيسية للمقاييس المستخدمة لتقييم الأنظمة المدمجة داخل المؤسسات والإدارة العامة.

غالبًا ما يشتمل النظام على آلية للتغذية المرتدة. تحدث التغذية الراجعة عندما يولد أحد مكونات النظام تدفق مخرجات يتم تغذيته مرة أخرى في النظام نفسه كتدفق إدخال ، ربما نتيجة لتحول إضافي.

تسمى الأنظمة القادرة على تعديل تدفقات المخرجات الخاصة بها بناءً على التغذية المرتدة أنظمة الدورة المغلقة. على سبيل المثال ، يصف نظام الدورة المغلقة الموضح في الشكل 2.2 تطوير سلسلة من الحملات التسويقية.



الشكل 2.1 التمثيل المجرد للنظام



الشكل 2.2 نظام تسويق الدورة المغلقة مع تأثيرات التغذية الراجعة

يتم جمع نتائج المبيعات لكل حملة وإتاحتها كمدخلات تغذية مرتدة وذلك لتصميم حملات ترويجية لاحقة.

فيما يتعلق بعملية صنع القرار ، التي سيتم وصف هيكلها في القسم التالي ، غالبًا ما يكون من الضروري تقييم أداء النظام. لهذا الغرض، من المناسب تصنيف **مقاييس التقييم** إلى فئتين رئيسيتين: الفعالية والكفاءة.

الفعالية Effectiveness. تعبر قياسات الفعالية عن مستوى مطابقة نظام معين للأهداف التي صُمم من أجلها. لذلك ، ترتبط مؤشرات الأداء المرتبطة بتدفقات مخرجات النظام ، مثل أحجام الإنتاج والمبيعات الأسبوعية والعائد لكل سهم.

الكفاءة Efficiency. تسلط قياسات الكفاءة الضوء على العلاقة بين تدفقات المدخلات التي يستخدمها النظام وتدفقات المخرجات المقابلة. ولذلك ترتبط قياسات الكفاءة بجودة عملية التحول. على سبيل المثال ، قد يعبرون عن مقدار الموارد اللازمة لتحقيق حجم مبيعات معين.

بشكل عام ، تشير مقاييس الفعالية إلى ما إذا كان يتم تنفيذ الإجراء الصحيح أم لا ، بينما توضح مقاييس الكفاءة ما إذا كان الإجراء يتم تنفيذه بأفضل طريقة ممكنة أم لا.

2.2 تمثيل عملية صنع القرار

من أجل بناء أنظمة دعم القرار فعالة ، نحتاج أولاً إلى وصف بعبارات عامة كيفية صياغة عملية صنع القرار. على وجه الخصوص ، نود أن نفهم الخطوات التي تقود الأفراد إلى اتخاذ

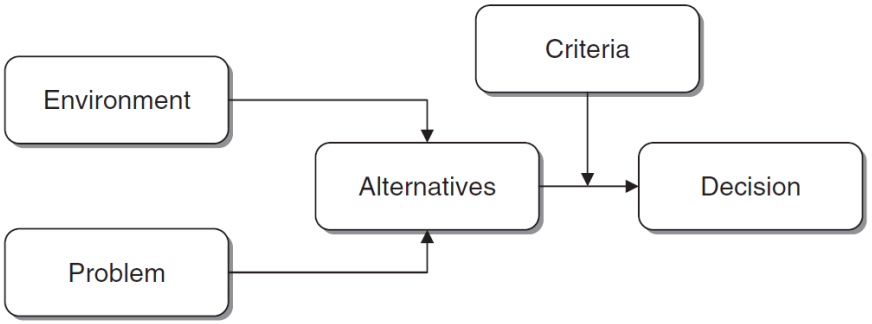
القرارات ومدى التأثير الذي تمارسه عليهم المواقف الذاتية لصانعي القرار والسياق المحدد الذي يتم فيه اتخاذ القرارات.

2.2.1 العقلانية وحل المشكلات

القرار هو اختيار من بين بدائل متعددة ، وعادة ما يتم بدرجة معقولة من العقلانية. يواجه كل فرد على أساس مستمر قرارات يمكن أن تكون أكثر أو أقل أهمية ، سواء في حياتهم الشخصية أو المهنية. في هذا القسم ، سوف نركز على القرارات التي يتخذها العاملون في مجال المعرفة في المؤسسات والمؤسسات العامة والخاصة. قد تتعلق هذه القرارات بوضع خطة إستراتيجية وبالتالي تتضمن خيارات استثمارية كبيرة ، وتعريف مبادرات التسويق وتوقعات المبيعات ذات الصلة ، وتصميم خطة إنتاج تسمح بتوظيف الموارد البشرية والتكنولوجية المتاحة بطريقة فعالة ومؤثرة.

تعد عملية صنع القرار جزءًا من موضوع أوسع يشار إليه عادةً باسم حل المشكلات ، والذي يشير إلى العملية التي يحاول الأفراد من خلالها سد الفجوة بين ظروف التشغيل الحالية للنظام (كما هي as is) والظروف الأفضل المفترض أن تكون تحقق في المستقبل (أن تكون to be). بشكل عام ، فإن انتقال النظام إلى الحالة المرغوبة يعني التغلب على بعض العقبات والتي ليس من السهل تحقيقها. وهذا يفرض على صانعي القرار ابتكار مجموعة من الخيارات البديلة الممكنة لتحقيق الهدف المنشود ، ثم اختيار قرار بناءً على مقارنة بين مزايا وعيوب كل خيار. ومن ثم ، يجب وضع القرار المختار موضع التنفيذ ومن ثم التحقق منه لتحديد ما إذا كان قد مكّن من تحقيق الأهداف المخطط لها. عندما لا يحدث هذا ، يتم إعادة النظر في المشكلة ، وفقًا لمنطق تعاودي.

يوضح الشكل 2.3 هيكل عملية حل المشكلات. تمثل البدائل الإجراءات الممكنة التي تهدف إلى حل المشكلة المحددة والمساعدة في تحقيق الهدف المخطط. في بعض الحالات ، قد يكون عدد البدائل التي يتم النظر فيها صغيرًا. في حالة وجود وكالة ائتمان يتعين عليها أن تقرر ما إذا كانت ستمنح قرضًا لمقدم الطلب أم لا ، يوجد خياران فقط ، وهما قبول الطلب ورفضه. في حالات أخرى ، يمكن أن يكون عدد البدائل كبيرًا جدًا أو حتى غير محدود. على سبيل المثال ، يتطلب تطوير الخطة اللوجيستية السنوية لشركة تصنيع الاختيار من بين عدد لا حصر له من الخيارات البديلة.



الشكل 2.3 التدفق المنطقي لعملية حل مشكلة

المعايير هي قياسات فعالية البدائل المختلفة والمتوافقة مع الأنواع المختلفة لأداء النظام الموضحة في الشكل 2.1. النهج العقلاني لصنع القرار يعني أن الخيار الذي يحقق أفضل معايير الأداء يتم اختياره من بين جميع البدائل الممكنة.

إلى جانب المعايير الاقتصادية ، التي تميل إلى السيادة في عملية صنع القرار داخل الشركات ، فمن الممكن تحديد العوامل الأخرى التي تؤثر على الاختيار العقلاني.

اقتصادية. العوامل الاقتصادية هي الأكثر تأثيرًا في عمليات صنع القرار ، وغالبًا ما تهدف إلى تقليل التكاليف أو تعظيم الأرباح. على سبيل المثال ، قد تكون الخطة اللوجستية السنوية مفضلة على الخطط البديلة إذا حققت انخفاضًا في إجمالي التكاليف.

اصطلاحية. يجب تجاهل الخيارات غير المجدية تقنيًا. على سبيل المثال ، لا يمكن اعتبار خطة الإنتاج التي تتجاوز السعة القصوى للمصنع خيارًا ممكنًا.

قانونية. تعني العقلانية القانونية أنه قبل تبني أي خيار ، يجب على صانعي القرار التحقق مما إذا كان متوافقًا مع التشريعات السارية في مجال التطبيق.

أخلاقية. إلى جانب الامتثال للقانون ، يجب أن يلتزم القرار بالمبادئ الأخلاقية والقواعد الاجتماعية للمجتمع الذي ينتمي إليه النظام.

إجرائية. يمكن اعتبار القرار مثاليًا من وجهة نظر اقتصادية وقانونية واجتماعية ، ولكنه قد يكون غير عملي بسبب القيود الثقافية للمنظمة من حيث الإجراءات السائدة والممارسات الشائعة.

سياسية. يجب على صانع القرار أيضًا تقييم النتائج السياسية لقرار معين بين الأفراد والإدارات والمنظمات.

يمكن تقسيم عملية تقييم البدائل إلى مرحلتين رئيسيتين ، كما هو موضح في الشكل 2.4: الاستبعاد والتقييم. أثناء مرحلة الاستبعاد ، يتم تطبيق قيود وقواعد التوافق على الإجراءات البديلة التي تم تحديدها في الأصل. ضمن عملية التقييم هذه ، سيتم إسقاط بعض البدائل من الدراسة ، بينما يمثل الباقي خيارات مجدية سيتم الترويج لها للتقييم. في مرحلة التقييم ، تتم مقارنة البدائل الممكنة مع بعضها البعض على أساس معايير الأداء ، من أجل تحديد القرار المفضل باعتباره أفضل فرصة.

2.2.2 عملية صنع القرار

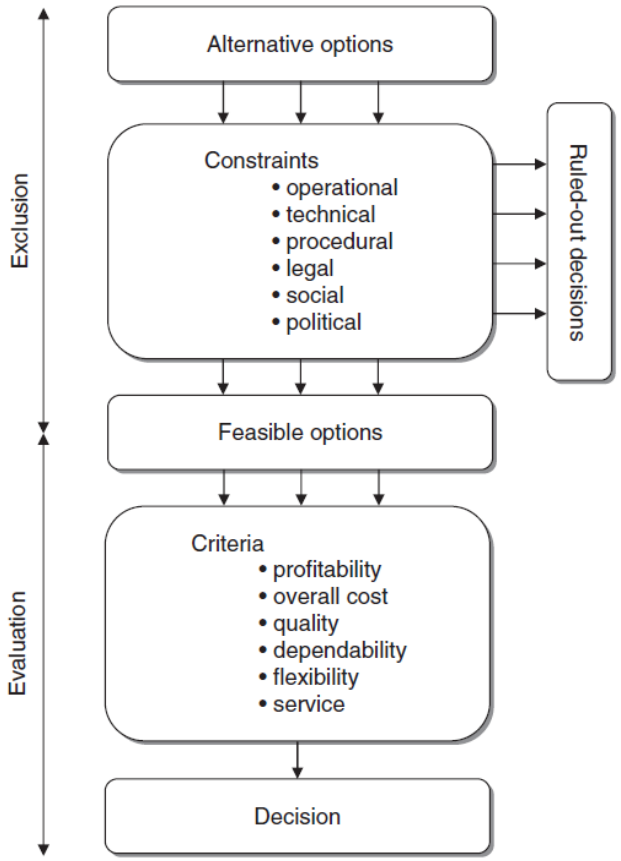
تم اقتراح تمثيل مقنع لعملية صنع القرار في أوائل الستينيات ، ولا يزال حتى اليوم مرجعًا منهجيًا رئيسيًا. يشتمل النموذج على ثلاث مراحل ، تسمى الذكاء والتصميم والاختيار. يوضح الشكل 2.5 نسخة موسعة من المخطط الأصلي ، والذي ينتج عن إدراج مرحلتين إضافيتين ، وهما التنفيذ والتحكم.

في مرحلة الذكاء ، تتمثل مهمة صانع القرار في تحديد المشكلة التي تظهر في النظام قيد الدراسة وتقييدها وتعريفها بوضوح. قد يسمح تحليل السياق وجميع المعلومات المتاحة لصانعي القرار بفهم سريع للإشارات والأعراض التي تشير إلى إجراء تصحيحي لتحسين أداء النظام. على سبيل المثال ، أثناء تنفيذ المشروع ، قد تتكون مرحلة الاستخبارات من مقارنة بين التقدم الحالي للأنشطة وخطة التطوير الأصلية. بشكل عام ، من المهم عدم الخلط بين المشكلة والأعراض.

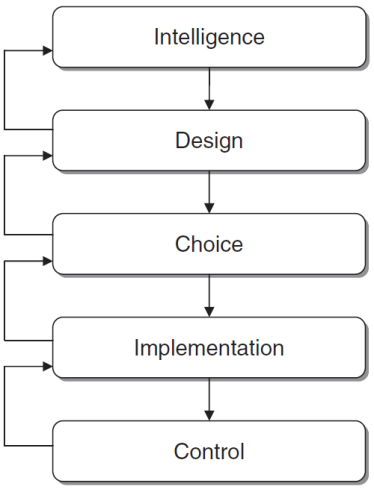
على سبيل المثال ، افترض أن بائع كتب التجارة الإلكترونية تلقى شكوى بخصوص تأخر تسليم طلب كتاب تم وضعه عبر الإنترنت. يمكن تفسير هذا الإزعاج على أنه مشكلة ويمكن معالجته من خلال ترتيب تسليم ثانٍ من خلال أولوية الشحن للتحايل على استياء العميل. من ناحية أخرى ، قد يكون هذا أحد أعراض مشكلة أوسع ، بسبب نقص الموظفين في قسم الشحن حيث من المحتمل أن تنشأ أخطاء بشرية تحت الضغط.

في مرحلة التصميم ، يجب تطوير وتخطيط الإجراءات التي تهدف إلى حل المشكلة المحددة. في هذا المستوى ، تلعب خبرة وإبداع صانعي القرار دورًا حاسمًا ، حيث يُطلب منهم ابتكار حلول قابلة للتطبيق تسمح في النهاية بتحقيق الهدف المقصود. عندما يكون عدد الإجراءات المتاحة صغيراً ، يمكن لصانعي القرار إجراء تعداد صريح للبدائل لتحديد أفضل الحلول. من ناحية أخرى

، إذا كان عدد البدائل كبيرًا جدًا ، أو حتى غير محدود ، فإن تحديدها يحدث بطريقة ضمنية ، عادةً من خلال وصف القواعد التي يجب أن تفي بها الإجراءات الممكنة. على سبيل المثال ، قد تُترجم هذه القواعد مباشرةً إلى قيود نموذج التحسين.



الشكل 2.4 الهيكل المنطقي لعملية صنع القرار الذكاء.



الشكل 2.5 مراحل عملية صنع القرار التصميم.

الأختيار. بمجرد تحديد الإجراءات البديلة ، من الضروري تقييمها على أساس معايير الأداء التي تعتبر مهمة. عادة ما تلعب النماذج الرياضية وطرق الحل المقابلة دورًا مهمًا خلال مرحلة الاختيار. على سبيل المثال ، تسمح نماذج وأساليب التحسين بإيجاد أفضل الحلول في المواقف المعقدة للغاية التي تتضمن عددًا لا يحصى من الحلول الممكنة أو حتى اللانهائية. من ناحية أخرى ، يمكن استخدام أشجار القرار للتعامل مع عمليات صنع القرار المتأثرة بالأحداث العشوائية.

تطبيق. عندما يتم اختيار أفضل بديل من قبل صانع القرار ، فإنه يتحول إلى إجراءات عن طريق خطة التنفيذ. وهذا ينطوي على تحديد المسؤوليات والأدوار لجميع المشاركين في خطة العمل.

مراقبة. بمجرد تنفيذ الإجراء ، من الضروري أخيرًا التحقق والتحقق من تلبية التوقعات الأصلية وأن آثار الإجراء تتطابق مع النوايا الأصلية. على وجه الخصوص ، يجب قياس الفروق بين قيم مؤشرات الأداء المحددة في مرحلة الاختيار والقيم التي لوحظت بالفعل في نهاية خطة التنفيذ. في نظام DSS المخطط بشكل مناسب ، تُترجم نتائج هذه التقييمات إلى خبرة ومعلومات ، والتي يتم نقلها بعد ذلك إلى مستودع البيانات لاستخدامها أثناء عمليات صنع القرار اللاحقة.

يمكن تلخيص الجوانب الأكثر صلة بعملية صنع القرار بإيجاز على النحو التالي.

- غالبًا ما يتم وضع القرارات من قبل مجموعة من الأفراد بدلاً من صانع قرار واحد.
- قد يكون عدد الإجراءات البديلة مرتفعًا جدًا وأحيانًا غير محدود.
- عادة ما تظهر آثار قرار معين لاحقًا ، وليس فورًا.
- غالبًا ما تكون القرارات المتخذة داخل مؤسسة أو مؤسسة عامة أو خاصة مترابطة وتحدد آثارًا واسعة النطاق. كل قرار له عواقب على العديد من الأفراد وعدة أجزاء من المنظمة.
- خلال عملية صنع القرار ، يُطلب من العاملين في مجال المعرفة الوصول إلى البيانات والمعلومات ، والعمل عليها بناءً على إطار مفاهيمي وتحليلي.
- تلعب التغذية الراجعة دورًا مهمًا في توفير المعلومات والمعرفة اللازمة لعمليات صنع القرار في المستقبل داخل منظمة معينة.
- في معظم الحالات ، تكون لعملية صنع القرار أهداف متعددة ، مع مؤشرات أداء مختلفة ، والتي قد تتعارض أيضًا مع بعضها البعض.

• يتم اتخاذ العديد من القرارات في سياق غامض وتنطوي على عوامل خطر.
يختلف مستوى الميل أو النفور من المخاطرة بشكل كبير بين الأفراد المختلفين.
• التجارب التي يتم إجراؤها في نظام العالم الحقيقي ، وفقًا لخطة التجربة والخطأ ، مكلفة للغاية ومحفوفة بالمخاطر بحيث لا يمكن استخدامها عمليًا في صنع القرار.
• تشير الديناميكيات التي تعمل فيها المؤسسة ، والتي تتأثر بشدة بضغط البيئة التنافسية ، إلى أن العاملين في مجال المعرفة بحاجة إلى معالجة المواقف واتخاذ القرارات بسرعة وفي الوقت المناسب.

2.2.3 أنواع القرارات

قد يكون تحديد تصنيف للقرارات مفيدًا أثناء تصميم DSS ، حيث من المحتمل أن عمليات صنع القرار ذات الخصائص المتشابهة يمكن أن تدعمها نفس مجموعة المنهجيات. يمكن تصنيف القرارات من حيث بعدين رئيسيين ، حسب طبيعتها ونطاقها. سيتم تقسيم كل بُعد إلى ثلاث فئات ، مما يعطي إجماليًا تسع مجموعات ممكنة ، كما هو موضح في الشكل 2.6.

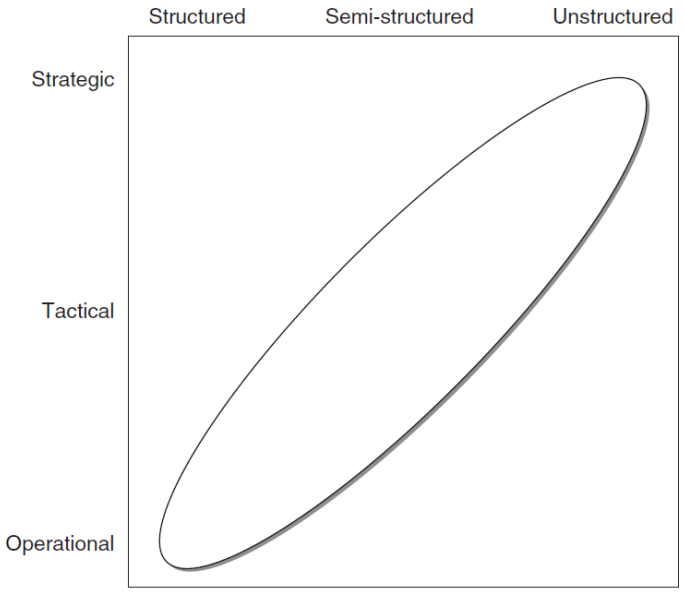
وفقًا لطبيعتها ، يمكن تصنيف القرارات على أنها منظمة أو غير منظمة أو شبه منظمة.

قرارات منظمة. يتم تنظيم القرار إذا كان يستند إلى إجراءات اتخاذ قرار محددة جيدًا ومتكررة. في معظم الحالات ، يمكن إرجاع القرارات المنظمة إلى خوارزمية ، والتي قد تكون أكثر أو أقل وضوحًا لصانعي القرار ، وبالتالي فهي مناسبة بشكل أفضل للأتمتة. وبشكل أكثر تحديدًا ، لدينا قرار منظم إذا كان من الممكن وصف تدفقات المدخلات وتدفقات المخرجات والتحولات التي يقوم بها النظام بوضوح في المراحل الثلاث للذكاء والتصميم والاختيار. في هذه الحالة ، سنقول أيضًا أن كل مرحلة مكونة يتم تنظيمها بدورها. في الواقع ، حتى القرارات التي تبدو منظمة بشكل كامل تتطلب في معظم الحالات التدخل المباشر من صانعي القرار للتعامل مع الأحداث غير المتوقعة ، الناجمة على سبيل المثال عن القيم غير العادية لبعض تدفقات المدخلات.

قرارات غير منظمة. يُقال إن القرار غير منظم إذا كانت المراحل الثلاث للذكاء والتصميم والاختيار غير منظمة أيضًا. وهذا يعني أنه لكل مرحلة يوجد عنصر واحد على الأقل في النظام (تدفقات المدخلات وتدفقات المخرجات وعمليات التحويل) لا يمكن وصفه بالتفصيل

واختزاله إلى تسلسل محدد مسبقًا من الخطوات. قد يحدث مثل هذا الحدث عندما يتم مواجهة عملية صنع القرار لأول مرة أو إذا كان نادرًا جدًا. في هذا النوع من القرارات ، يكون دور العاملين في مجال المعرفة أساسيًا ، وقد توفر أنظمة ذكاء الأعمال الدعم لصانعي القرار من خلال الوصول إلى المعلومات في الوقت المناسب ومتعدد الاستخدامات.

القرارات شبه المنظمة. يكون القرار شبه منظم عندما يتم تنظيم بعض المراحل والبعض الآخر ليس كذلك. معظم القرارات التي يواجهها العاملون في مجال المعرفة في إدارة المؤسسات أو المنظمات العامة أو الخاصة هي شبه منظمة. بالتالي ، يمكنهم الاستفادة من DSS وبيئة ذكاء الأعمال بطريقتين أساسيتين. بالنسبة للمراحل غير المنظمة من عملية صنع القرار ، قد تقدم أدوات ذكاء الأعمال نوعًا من الدعم السلبي الذي يترجم إلى الوصول إلى المعلومات في الوقت المناسب ومتعدد الاستخدامات. بالنسبة للمراحل المنظمة ، من الممكن توفير شكل نشط من الدعم من خلال النماذج والخوارزميات الرياضية التي تسمح بأتمتة أجزاء كبيرة من عملية صنع القرار.



2.6 تصنيف للقرارات

في بعض الأحيان قد تنشأ المواقف التي لا يمكن فيها تحديد طبيعة القرار بسهولة بشكل لا لبس فيه. عند مواجهة نفس المشكلة ، مثل تحديد سعر بيع أحد المنتجات ، قد يأتي صانعو القرار المختلفون العاملون في مؤسسات مختلفة بخيارات مختلفة. على سبيل المثال ، قد يعتقد صانع القرار الأول أنه يمكن الحصول على أفضل سعر بيع من خلال مقارنة منحنيات مرونة السعر

والطلب. نتيجة لذلك ، قد يعتبر صانع القرار هذا أن مرحلة الاختيار في عملية صنع القرار منظمة. على النقيض من ذلك ، قد يعتقد صانع القرار الثاني أن منحنى المرونة لا يعكس جميع العوامل التي تؤثر على استجابة السوق لتقلبات الأسعار لأن بعض هذه العناصر لا يمكن قياسها كمياً.

بالنسبة لهذا الفرد ، تبين أن مرحلة الاختيار غير منظمة أو شبه منظمة. تصف الأمثلة 2.1 و 2.2 و 2.3 قرارات منظمة وشبه منظمة وغير منظمة ، على التوالي.

مثال 2.1 - قرار منظم. ينتج مصنع الورق أوراقًا ورقية لمستودع الشركة بأحجام قياسية مختلفة يتم قصها بشكل متسلسل حسب الحجم للعملاء. على وجه التحديد ، يقوم العملاء بإرسال الطلبات من حيث نوع الورق والكمية والحجم. الأحجام المحددة في الطلبات عادة ما تكون أصغر من الأحجام القياسية ويجب قصها منها.

لذلك ، فإن مصنع الورق مجبر على النظر في أفضل طريقة للجمع بين الأحجام المطلوبة للوفاء بالطلبات وقطعها من الأحجام القياسية لتقليل نفايات الورق. هذا القرار شائع في العديد من الصناعات (الورق والألمنيوم والخشب والصلب والزجاج والنسيج) ويمكن دعمه جيدًا من خلال نماذج التحسين. ومع ذلك ، حتى فيما يتعلق بهذه القرارات المنظمة ، قد تتطلب ظروف معينة وقيم إدخال محددة تدخل صانع القرار لتعديل الخطط التي تم الحصول عليها عن طريق نماذج التحسين. على سبيل المثال ، قد ترغب الشركة في تفضيل طلب محدد للعميل يعتبر استراتيجيًا ، وإدخال مسار معالجة سريعة في خطة القطع ، حتى لو كان هذا قد ينطوي على المزيد من المواد المهدرة أثناء مرحلة القطع.

مثال 2.2 - قرار شبه منظم. يحتاج مدير الخدمات اللوجستية لشركة تصنيع إلى تطوير خطة سنوية. تحدد الخطة اللوجستية تخصيص أحجام الإنتاج المتوقعة لكل مصنع لمناطق السوق المختلفة ، وشراء المواد من كل مورد مع الأحجام ذات الصلة وأوقات التسليم ، ودرجات الإنتاج لكل مرحلة تصنيع ، ومستويات مخزون التجميعات الفرعية والعناصر النهائية ، وتوزيع العناصر النهائية على مناطق السوق.

هذه القرارات لها تأثير اقتصادي وتنظيمي كبير قد يستفيد بشكل كبير من اعتماد DSS على أساس نماذج التحسين واسعة النطاق. ومع ذلك ، فمن المحتمل أنه في الوضع الحقيقي ، تُترك بعض العناصر لتقدير صانعي القرار ، الذين قد يفضلون خطة لوجستية معينة على أخرى ، حتى

لو كانت تنطوي على تكاليف أعلى إلى حد ما مقارنة بالخطة المثلى التي اقترحها النموذج. على سبيل المثال ، قد يكون من المناسب الحفاظ على الإمداد غير المتغير للأجزاء المشتراة من مورد معين والذي يعتبر إستراتيجياً للمستقبل على الرغم من أن هذا المورد أقل قدرة على المنافسة من غيره ، والتي يفضلها نموذج التحسين بدلاً من ذلك من حيث التكلفة الدنيا.

مثال 2.3 - قرار غير منظم. ضع في اعتبارك مؤسسة هدفها استحواذ عدائي من خلال عرض عام مقدم من قبل منافس مباشر. هناك العديد من القرارات والإجراءات الدفاعية المحتملة التي تعتمد بشدة على السياق الذي تعمل فيه المؤسسة ويتم تقديم العرض. من الصعب تصور وصف منهجي لعملية اتخاذ القرار يمكن إعادة إنتاجه لاحقًا في حالات أخرى مماثلة.

يتضح من الأمثلة المذكورة أعلاه أن طبيعة عملية اتخاذ القرار تعتمد على العديد من العوامل ، بما في ذلك: • خصائص المنظمة التي يتم وضع النظام فيها. • المواقف الذاتية لمتخذي القرار. • توافر منهجيات مناسبة لحل المشكلات. • توافر أدوات فعالة لدعم القرار.

اعتماداً على نطاقها ، يمكن تصنيف القرارات على أنها استراتيجية وتكتيكية وتشغيلية.

القرارات الإستراتيجية. تعتبر القرارات إستراتيجية عندما تؤثر على المنظمة بأكملها أو على الأقل جزء كبير منها لفترة طويلة من الزمن. تؤثر القرارات الإستراتيجية بقوة على الأهداف والسياسات العامة للمؤسسة.

نتيجة لذلك ، يتم اتخاذ القرارات الإستراتيجية على مستوى تنظيمي أعلى ، عادة من قبل الإدارة العليا للشركة.

قرارات تكتيكية. تؤثر القرارات التكتيكية على أجزاء فقط من المؤسسة وعادةً ما تقتصر على قسم واحد. الفترة الزمنية محدودة بأفق متوسط المدى ، وعادة ما يصل إلى عام. تضع القرارات التكتيكية نفسها في السياق الذي تحدده القرارات الإستراتيجية. في التسلسل الهرمي للشركة ، يتم اتخاذ القرارات التكتيكية من قبل المديرين المتوسطين ، مثل رؤساء أقسام الشركة.

القرارات التشغيلية. تشير القرارات التشغيلية إلى أنشطة محددة يتم تنفيذها داخل منظمة ويكون لها تأثير متواضع على المستقبل. يتم تأطير القرارات التشغيلية ضمن العناصر والشروط التي تحددها القرارات الاستراتيجية والتكتيكية. لذلك ، عادة ما يتم إجراؤها على مستوى تنظيمي

أدنى ، من قبل عمال المعرفة المسؤولين عن نشاط أو مهمة واحدة مثل رؤساء الأقسام الفرعية ، ورش العمل ، ورؤساء المكاتب الخلفية.

ستتغير خصائص المعلومات المطلوبة في عملية صنع القرار اعتمادًا على نطاق القرارات التي سيتم دعمها ، وبالتالي أيضًا سيختلف اتجاه DSS وفقًا لذلك. يوضح الشكل 2.7 الاختلافات في خصائص المعلومات مع تغير نطاق القرارات. يمكن استخدام المخطط كأداة تقييم عند تصميم DSS: بمجرد تحديد نطاق القرارات التي يهدف النظام من أجلها ، يمكن استخدام المخطط لتحديد ما إذا كانت عملية صنع القرار مدعومة بشكل كافٍ بالمعلومات الصحيحة .

على الرغم من أن الطبيعة والنطاق غير مترابطين تمامًا ، إلا أن معظم قرارات العالم الحقيقي تقع ضمن القطع الناقص الموضح في الشكل 2.6: معظم القرارات الإستراتيجية غير منظمة ، في حين أن معظم القرارات التشغيلية منظمة ومعظم القرارات التكتيكية شبه هيكلية ملاح. هذه الملاحظة التجريبية مفيدة عند التحديد المسبق لخصائص DSS لتسهيل عملية صنع القرار ذات الطبيعة والنطاق المحددين.

	Operational	Tactical	Strategic
Accuracy	High	⟷	Low
Level of detail	Detailed	⟷	Aggregate
Time horizon	Present	⟷	Future
Frequency of use	High	⟷	Low
Source	Internal	⟷	External
Scope of information	Quantitative	⟷	Qualitative
Nature of information	Narrow	⟷	Wide
Age of information	Present	⟷	Past

الشكل 2.7 خصائص المعلومات من حيث نطاق القرارات

2.2.4 مناهج عملية صنع القرار

كما لوحظ أعلاه ، فإن التوجه الذاتي لصانعي القرار عبر المنظمة يؤثر بشدة على هيكل عملية صنع القرار.

في هذا القسم سنراجع المناهج الرئيسية التي تسود في إدارة المنظمات المعقدة ، وندرس الآثار عند تصميم DSS.

يجب أن يتم التمييز الأولي بين النهج العقلاني والنهج السياسي التنظيمي.

النهج عقلاني. عندما يتم اتباع نهج عقلاني ، يأخذ صانع القرار في الاعتبار العوامل الرئيسية ، مثل الاقتصادية والتقنية والقانونية والأخلاقية والإجرائية والسياسية ، وكذلك تحديد معايير التقييم لتقييم الخيارات المختلفة ثم اختيار القرار الأفضل. في هذا السياق ، قد يساعد نظام DSS بطريقة سلبية ، من خلال الوصول إلى المعلومات في الوقت المناسب ومتعدد الاستخدامات ، وبطريقة نشطة ، من خلال استخدام النماذج الرياضية لصنع القرار.

النهج السياسي التنظيمي. عندما يتم اتباع نهج سياسي تنظيمي ، يتقدم صانع القرار بطريقة غريزية أكثر وأقل منهجية. لا تستند القرارات إلى بدائل محددة بوضوح ومعايير الاختيار. نتيجة لذلك ، يمكن لـ DSS المساعدة فقط بطريقة سلبية ، مما يوفر وصولاً متنوعًا وفي الوقت المناسب إلى المعلومات. قد يكون مفيدًا أيضًا أثناء المناقشات والمفاوضات في عمليات صنع القرار تلك التي تشمل جهات فاعلة متعددة ، مثل المديرين العاملين في أقسام مختلفة.

ضمن النهج العقلاني يمكننا التمييز بين طريقتين بديلتين تؤثر فيهما عملية صنع القرار الفعلية على القرارات: العقلانية المطلقة والعقلانية المحدودة.

العقلانية المطلقة. يشير مصطلح "العقلانية المطلقة" إلى عملية اتخاذ القرار التي يمكن من خلالها تقليل مؤشرات الأداء المتعددة إلى معيار واحد ، مما يفسح المجال بشكل طبيعي لنموذج التحسين. على سبيل المثال ، قد يكون مدير الإنتاج الذي يتعين عليه وضع خطة لوجستية متوسطة الأجل قادرًا على تحويل جميع مؤشرات الأداء إلى وحدات نقدية ، وبالتالي اشتقاق الحل لاحقًا بأقل تكلفة. وهذا يعني أن المؤشرات غير النقدية ، مثل أحجام المخزون أو عدد أيام التأخير في التعامل مع أمر معين ، يجب أن تتحول إلى وحدات قياس نقدي. من منظور منهجي ، هذا يعني أن مشكلة التحسين متعددة الأهداف تتحول إلى مشكلة ذات هدف واحد من

خلال التعبير عن جميع العوامل ذات الصلة في وحدة قياس مشتركة تسمح بإضافة الأهداف غير المتجانسة معًا.

العقلانية المحدودة. تحدث العقلانية المقيدة عندما لا يكون من الممكن تقليل معايير متعددة إلى هدف واحد ، بحيث يعتبر صانع القرار أن الخيار يكون مرضيًا عندما تكون مؤشرات الأداء المقابلة أعلى أو أقل من قيم العتبة المحددة مسبقًا. على سبيل المثال ، تكون خطة الإنتاج مقبولة إذا كانت تكلفتها منخفضة بما فيه الكفاية ، وكميات المخزون ضمن حد معين ، ووقت الخدمة أقل من توقعات العملاء. لذلك ، فإن مفهوم العقلانية المحدودة يجسد الخيارات العقلانية المقيدة بحدود المعرفة والقدرة المعرفية.

تهدف معظم عمليات صنع القرار التي تحدث داخل المؤسسات والإدارة العامة إلى اتخاذ قرار يبدو مقبولًا فيما يتعلق بمعايير التقييم المتعددة ، وبالتالي من المرجح أن تحدث عمليات القرار القائمة على العقلانية المحدودة في الممارسة العملية.

2.3 تطور نظم المعلومات

تجمع أنظمة دعم القرار بين البيانات والنماذج الرياضية لمساعدة صانعي القرار في عملهم. إلى حد ما تظهر صلات مع أنظمة المعلومات للمؤسسة. ومن ثم ، يجدر وصف الجدول الزمني لتطور نظم المعلومات في هذا القسم ، لأن هذا يسلط الضوء على كيفية تطور معالجة البيانات واستخدامها داخل الشركات.

ظهرت أجهزة الكمبيوتر الرقمية في أواخر الأربعينيات ، وسرعان ما بدأ تطبيقها في بيئة الأعمال. شهدت العقود الأولى اندفاعًا نحو تطوير تكنولوجيا المعلومات ، عادةً تحت شعار معالجة البيانات. وقد تميزوا بالانتشار الواسع للتطبيقات التي حققت زيادة في الكفاءة من خلال أتمتة العمليات الروتينية داخل الشركات ، وخاصة في الإدارة والإنتاج والبحث والتطوير.

في السبعينيات من القرن الماضي ، بدأت تظهر داخل الشركات احتياجات معقدة بشكل متزايد لابتكار تطبيقات برمجية ، تسمى أنظمة المعلومات الإدارية (MIS) ، من أجل تسهيل الوصول إلى المعلومات المفيدة وفي الوقت المناسب لصانعي القرار. ومع ذلك ، أعاقت حالة تقنيات المعلومات في ذلك الوقت محاولات تطوير مثل هذه الأنظمة. افتقرت أجهزة الكمبيوتر

المركزية في تلك الأيام إلى إمكانات التصور الرسومي ، وتواصلت مع المستخدمين من خلال المحطات الطرفية للكمبيوتر والطابعات النقطية. تكمن صعوبة أخرى في الهيكل التنظيمي للشركات ، استنادًا إلى قسم نظم المعلومات شديد المركزية ، مما يؤدي عادةً إلى تأخيرات زمنية طويلة جدًا ومحبطة في تنفيذ التغييرات أو التمديدات للتطبيقات المتاحة.

منذ أواخر الثمانينيات ، كان لإدخال أجهزة الكمبيوتر الشخصية مع أنظمة التشغيل التي تتميز بواجهات الرسوم وأجهزة التأشير ، مثل الماوس أو القلم البصري ، نتيجتان رئيسيتان. فمن ناحية ، أصبح من الممكن تنفيذ تطبيقات قادرة على التفاعلات المعقدة والعرض الرسومي للنتائج ، وهو شرط أساسي لتزويد صانعي القرار بأدوات دعم مفيدة حقًا. من ناحية أخرى ، يمكن للعاملين في مجال المعرفة الاعتماد على أدوات المعالجة المستقلة مما جعلهم مستقلين إلى حد كبير عن قسم نظم معلومات الشركة ، وبالتالي تجنب التأخير الزمني المذكور أعلاه في الوصول إلى البيانات. وقد أدى ذلك إلى قيام أكثر العاملين في مجال المعرفة نشاطًا بإنشاء قواعد بيانات محلية وتطوير نماذج محاكاة ، على سبيل المثال عن طريق جداول البيانات ، والتي يمكن اعتبارها أسلافًا حقيقيين لهياكل ذكاء الأعمال الحالية. كانت الزيادة في قدرات المعالجة المستقلة التي يحتفظ بها المستخدمون ، والتي يشار إليها عادةً باسم حوسبة المستخدم النهائي ، عاملاً تمكينيًا حاسمًا للتطورات المستقبلية ، حيث ساعدت في تحديد أهمية أقسام نظم المعلومات.

وفي الوقت نفسه ، تم أيضًا تقديم المفهوم الأولي لنظام دعم القرار ، والذي سيتم وصف معناه الدقيق في القسم التالي. سلطت التطورات اللاحقة الضوء على أنواع جديدة من التطبيقات والبنى: تم تقديم أنظمة المعلومات التنفيذية وأنظمة المعلومات الاستراتيجية لأول مرة في أواخر الثمانينيات لدعم المديرين التنفيذيين في عملية صنع القرار.

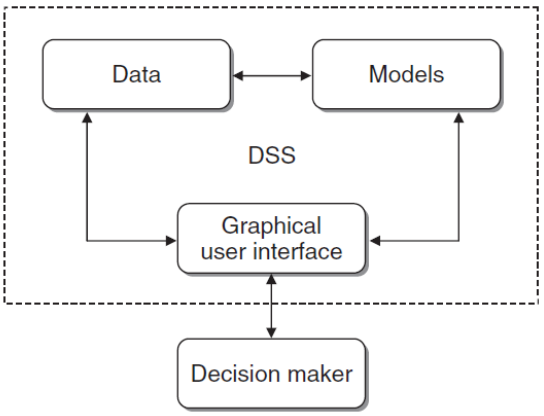
كانت هذه الأنظمة مخصصة لعمليات صنع القرار غير المنظمة ، وبالتالي تمثل أنظمة الدعم السلبي الموجهة نحو الوصول السريع والسهل إلى المعلومات. منذ أوائل التسعينيات ، بدأ اعتماد معماريات الشبكات وأنظمة المعلومات الموزعة القائمة على نماذج حوسبة العميل والخادم على نطاق واسع.

علاوة على ذلك ، نشأت الحاجة إلى الفصل المنطقي والمادي بين قواعد البيانات المخصصة لأنظمة DSS من أنظمة المعلومات التشغيلية. أدى ذلك إلى ظهور مفاهيم مستودعات البيانات وقطارات البيانات ، والتي سيتم وصفها في الفصل 3.

أخيرًا ، في نهاية التسعينيات ، بدأ استخدام مصطلح ذكاء الأعمال لمعالجة البنية التي تحتوي على DSS والمنهجيات والنماذج التحليلية المستخدمة لتحويل البيانات إلى معلومات ومعرفة مفيدة لصانعي القرار ، كما تمت مناقشته في الفصل 1.

2.4 تعريف نظام دعم القرار

منذ أواخر الثمانينيات ، تم تعريف نظام دعم القرار على أنه نظام كمبيوتر تفاعلي يساعد صانعي القرار على الجمع بين البيانات والنماذج لحل المشكلات شبه المنظمة وغير المنظمة. يستلزم هذا التعريف العناصر الرئيسية الثلاثة لـ DSS الموضحة في الشكل 2.8: قاعدة بيانات ، ومستودع للنماذج الرياضية ووحدة للتعامل مع الحوار بين النظام والمستخدمين. ومن ثم فهو يسلط الضوء على دور أنظمة دعم القرار كنقطة محورية لاتجاهات التطور في مجالين متميزين: من ناحية ، معالجة البيانات وتقنيات المعلومات ؛ ومن ناحية أخرى ، التخصصات التي تتناول دراسة النماذج والأساليب الرياضية ، مثل بحوث العمليات والإحصاءات.



الشكل 2.8 هيكل نظام دعم القرار

في الواقع ، على الرغم من التحسينات الهامة التي تم تحقيقها في كلا المجالين ، كان التنفيذ الفعلي للتطبيقات التي يمكن استخدامها من قبل العاملين في مجال المعرفة مزعجًا. كانت تكنولوجيا المعلومات في أوائل السبعينيات ، كما نوقش في القسم السابق ، تتألف بشكل أساسي من تطبيقات للمحاسبة والإدارة ، وقد اقتصرت قدراتها على معالجة كميات كبيرة من المعاملات وإنتاج تقارير موجزة. يشير الفشل الجزئي لأنظمة المعلومات الإدارية إلى أنه حتى الهدف

الأساسي المتمثل في الوصول إلى المعلومات في الوقت المناسب ومتعدد الاستخدامات لم يكن في المتناول.

من ناحية أخرى ، على الرغم من أن النماذج الرياضية أصبحت أكثر مرونة من خلال اللجوء إلى تقنيات تفاعل المستخدم المبتكرة ، إلا أنها كانت لا تزال تعتمد في الغالب على الأساليب التنظيمية ، ومناسبة للقرارات التشغيلية التي تتميز بأهداف فريدة بدلاً من القرارات التكتيكية والاستراتيجية ، وعادةً ما تكون أقل تنظيماً.

كما لاحظنا ، فإن عمليات اتخاذ القرار شبه المنظمة وغير المهيكلة تتأثر بشدة بالآراء الذاتية ، بحيث يتعين على العاملين في مجال المعرفة تحليل مؤشرات الأداء المتعددة بشكل مباشر والتي سيكون من الصعب تقليلها إلى معيار واحد لاتخاذ القرار.

يجدر تسليط الضوء على الميزات ذات الصلة لـ DSS من أجل تحديد التعريف الوارد أعلاه وفهم دوره بشكل أفضل.

فعالية. يجب أن تساعد أنظمة دعم القرار العاملين في مجال المعرفة على الوصول إلى قرارات أكثر فعالية. في هذا الصدد ، فهي عنصر أساسي في معماريات ذكاء الأعمال. لاحظ أن هذا لا يعني بالضرورة زيادة الكفاءة في عملية صنع القرار. في الواقع ، قد يستلزم اعتماد DSS تحليلًا أكثر دقة وبالتالي يتطلب استثمارًا أكبر للوقت من قبل صانعي القرار. ومع ذلك ، فإن الجهد الأكبر المطلوب سينتج عنه عادة قرارات أفضل.

النماذج الرياضية. من أجل تحقيق قرارات أكثر فاعلية ، يستخدم نظام دعم القرار نماذج رياضية ، مستعارة من تخصصات مثل أبحاث العمليات والإحصاءات ، والتي يتم تطبيقها على البيانات الموجودة في مستودعات البيانات وسوق البيانات. يعد استخدام النماذج التحليلية لتحويل البيانات إلى معرفة وتقديم الدعم النشط هو السمة الرئيسية التي تميز DSS عن نظام المعلومات البسيط.

الاندماج في عملية صنع القرار. يجب أن تقدم DSS المساعدة لأنواع مختلفة من العاملين في مجال المعرفة ، ضمن نفس مجال التطبيق ، لا سيما فيما يتعلق بعمليات اتخاذ القرار شبه المنظمة وغير المنظمة ، سواء ذات الطبيعة الفردية أو الجماعية. علاوة على ذلك ، فإن DSS مخصص لعمليات صنع القرار التي تكون استراتيجية وتكتيكية وتشغيلية في نطاقها.

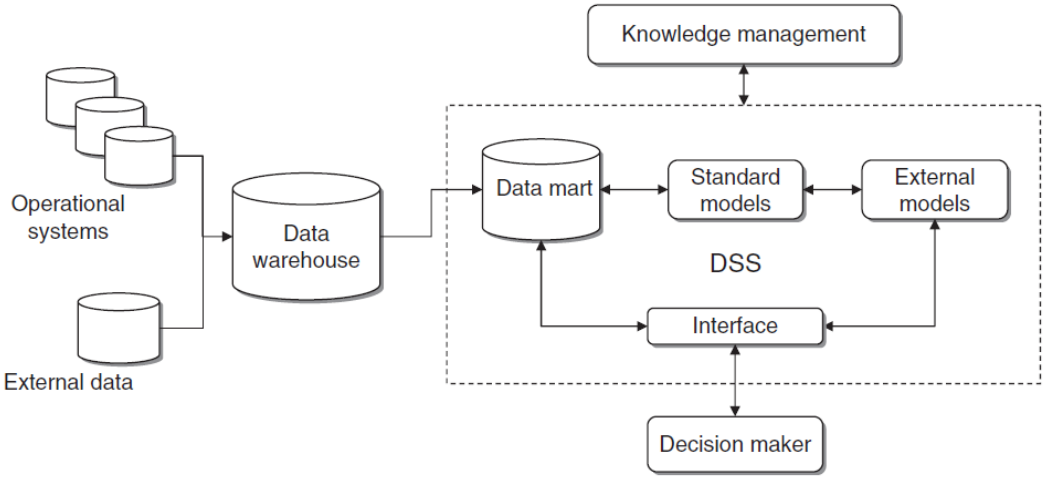
علاوة على ذلك ، يجب أن تتاح لصانعي القرار فرصة دمج تفضيلاتهم وكفاءاتهم في DSS ، وتكييفها مع احتياجاتهم بدلاً من القبول السلبي لما ينتج عنها. بهذه الطريقة ، قد تأخذ DSS تدريجياً دور مكون رئيسي في منهجية حل المشكلات التي يتبناها صانعو القرار ، مما يتيح أسلوب صنع القرار الاستباقي والمدرك ، بدلاً من الموقف التفاعلي والاستثنائي ، من أجل توقع أي ظاهرة ديناميكية سريعة التطور.

الدور التنظيمي. في العديد من المواقف ، يعمل مستخدمو DSS على مستويات هرمية مختلفة داخل المؤسسة ، ويميل DSS إلى تشجيع التواصل بين مختلف أجزاء المؤسسة. من خلال توفير الدعم لعمليات اتخاذ القرار المتسلسلة والمترابطة ، يمكن لنظام دعم القرار تتبع التحليل والمعلومات التي أدت إلى قرار محدد.

المرونة. يجب أن تكون DSS مرنة وقابلة للتكيف من أجل دمج التغييرات المطلوبة لتعكس التعديلات في البيئة أو في عملية صنع القرار. علاوة على ذلك ، يجب أن يكون سهل الاستخدام ، مع طرق تفاعل سهلة الاستخدام وبديهية ورسومات عالية الجودة لعرض المعلومات المستخرجة أو المنتجة. أصبح من الشائع بشكل متزايد أن تتميز أنظمة دعم القرار بواجهة متصفح ويب للتواصل مع المستخدمين.

تم تمديد هيكل DSS الموضح في الشكل 2.8 في الشكل 2.9 ليشمل بعض المكونات الجديدة.

إدارة البيانات. تشتمل وحدة إدارة البيانات على قاعدة بيانات مصممة لاحتواء البيانات التي تتطلبها عمليات صنع القرار التي يتم توجيه نظام دعم القرار إليها. في معظم التطبيقات ، تكون قاعدة البيانات عبارة عن سوق بيانات ، كما سنرى في الفصل 3. مع الأخذ في الاعتبار بنيات ذكاء الأعمال الحالية ، ترتبط وحدة إدارة البيانات في DSS عادةً بمخزن بيانات الشركة ، كما هو موضح في الفصل 3 ، والذي يمثل المستودع الرئيسي للبيانات المتاحة لتطوير تحليل ذكاء الأعمال.



الشكل 2.9 هيكل موسع لنظام دعم القرار

إدارة النموذج. توفر وحدة إدارة النموذج للمستخدمين النهائيين مجموعة من النماذج الرياضية المستمدة من بحوث العمليات والإحصاءات والتحليل المالي. عادة ما تكون هذه نماذج بسيطة نسبيًا تسمح بإجراء التحقيقات التحليلية التي تكون مفيدة جدًا أثناء عملية صنع القرار. لتوضيح الدور الذي تلعبه وحدة إدارة النموذج ، يمكن للمرء التفكير في الوظائف التحليلية التي توفرها جداول البيانات الحالية. وتشمل هذه نماذج التحسين البسيطة ونماذج التحليل المالي والاكتواري والوظائف الإحصائية. علاوة على ذلك ، تساعد وحدة إدارة النموذج أنشطة العاملين في مجال المعرفة عن طريق اللغات عالية المستوى لتطوير نماذج مخصصة. في تطبيقات معينة ، قد يتم دمج هذه الوحدة مع نماذج أكثر تعقيدًا ، يشار إليها بالنماذج الخارجية في الشكل 2.9 ، والتي تم إنشاؤها لتنفيذ مهام تحليل محددة. على سبيل المثال ، يندرج نموذج التحسين واسع النطاق المصمم لتطوير الخطة اللوجستية السنوية لشركة تصنيع ضمن هذه الفئة.

التفاعلات. في معظم التطبيقات ، يستخدم العاملون في مجال المعرفة DSS بشكل تفاعلي لإجراء تحليلاتهم. من المتوقع أن تتلقى الوحدة المسؤولة عن هذه التفاعلات بيانات الإدخال من المستخدمين بأسهل الطرق وأكثرها بديهية ، عادةً من خلال الواجهة الرسومية لمتصفح الويب ، ثم إرجاع المعلومات المستخرجة والمعرفة الناتجة عن النظام في رسم بياني مناسب شكل.

إدارة المعرفة. ترتبط وحدة إدارة المعرفة أيضًا بالنظام المتكامل لإدارة المعرفة في الشركة. يسمح لصانعي القرار بالاستفادة من الأشكال المختلفة للمعرفة الجماعية ، غير المنظمة عادة ، والتي تمثل ثقافة الشركة.

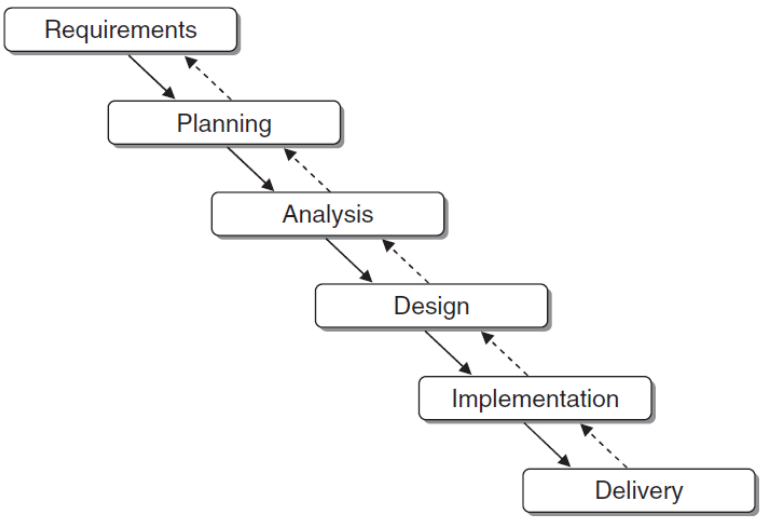
يختتم هذا القسم بملخص للمزايا المحتملة الرئيسية الناتجة عن اعتماد نظام دعم القرار:

- زيادة في عدد البدائل أو الخيارات التي يتم النظر فيها ؛
- زيادة عدد القرارات الفعالة التي يتم وضعها.
- وعي أكبر وفهم أعمق للمجال الذي تم تحليله والمشكلات التي تم بحثها.
- إمكانية تنفيذ تحليلات السيناريو وماذا لو من خلال تغيير الفرضيات والمعاملات الخاصة بالنماذج الرياضية.
- تحسين القدرة على الاستجابة بسرعة للأحداث غير المتوقعة والمواقف غير المتوقعة ؛
- استغلال القيمة المضافة للبيانات المتاحة.
- تحسين الاتصال والتنسيق بين الأفراد والإدارات التنظيمية.
- تطوير العمل الجماعي بشكل أكثر فعالية.
- زيادة موثوقية آليات التحكم ، بسبب زيادة الوضوح في عملية اتخاذ القرار.

2.5 تطوير نظام دعم القرار

في هذا القسم سوف نصف مراحل تطوير DSS. على عكس تطبيقات البرامج الأخرى ، مثل أنظمة المعلومات وأدوات التشغيل الآلي للمكاتب ، لا تتوفر أنظمة DSS عادةً كبرامج قياسية. سهلت بيئات التحليل متعددة الأبعاد الوصول إلى وظائف ذكاء الأعمال المنفعلة ووحدت معاييرها. ومع ذلك ، من أجل تطوير معظم أنظمة دعم القرار ، لا يزال هناك حاجة إلى مشروع معين.

يوضح الشكل 2.10 الخطوات الرئيسية في تطوير DSS. يتم عرض التدفق المنطقي للأنشطة بواسطة الأسهم الصلبة. تشير الأسهم المنقطة في الاتجاه المعاكس إلى مراجعات لمرحلة واحدة أو أكثر والتي قد تصبح ضرورية أثناء تطوير النظام ، من خلال آلية التغذية الراجعة. نصف الآن بالتفصيل كيف يتم تنفيذ كل مرحلة.



الشكل 2.10 مراحل تطوير تخطيط نظام دعم القرار.

التخطيط . الغرض الرئيسي من مرحلة التخطيط هو فهم الاحتياجات والفرص ، التي تتميز أحيانًا بإشارات ضعيفة ، وترجمتها إلى مشروع ثم لاحقًا إلى DSS ناجح. عادة ما يتضمن التخطيط دراسة جدوى لمعالجة السؤال: لماذا نرغب في تطوير DSS؟ أثناء تحليل الجدوى ، يتم تحديد الأهداف العامة والمحددة للنظام والمستفيدين والفوائد المحتملة وأوقات التنفيذ والتكاليف. ليس من السهل تحديد فوائد DSS. كما لوحظ بالفعل ، فإن الميزة الرئيسية لمعظم أنظمة دعم القرار لا تتمثل في زيادة الفعالية بقدر ما هي تحسين جودة القرار ، وهو أمر يصعب التنبؤ به مسبقًا. تصبح العناصر المطلوبة لتقرير ما إذا كان ينبغي تنفيذ النظام أم لا متوفرة في نهاية دراسة الجدوى. سيؤدي القرار السلبي إلى وقف تطوير المشروع ، على الرغم من أنه بالطبع قد يتم إعادة النظر في هذا في وقت لاحق. إذا قرر المرء المضي قدمًا في النظام ، فيجب أن يتبع مرحلة التخطيط تحديد الأنشطة والمهام والمسؤوليات ومراحل التطوير ، والتي يجب استخدام منهجيات إدارة المشروع الكلاسيكية.

التحليل. في مرحلة التحليل ، من الضروري تحديد وظائف DSS التي سيتم تطويرها بالتفصيل ، من خلال تطوير وتفصيل الاستنتاجات الأولية التي تحققت أثناء دراسة الجدوى. لذلك يجب الإجابة على السؤال التالي: ما الذي يجب أن تنجزه DSS ، ومن سيستخدمها ، ومتى وكيف؟ لتقديم إجابة ، من الضروري تحليل عمليات اتخاذ القرار التي سيتم دعمها ، لمحاولة فهم شامل لجميع العلاقات المتبادلة الموجودة بين المشكلات التي تمت معالجتها والبيئة المحيطة. يجب تقييم الآثار التنظيمية التي تحددها DSS. يتضمن التحليل أيضًا رسم خرائط لعمليات القرار الفعلية

وتخيل الشكل الذي ستبدو عليه العمليات الجديدة بمجرد وضع DSS في مكانه. أخيرًا ، من الضروري استكشاف البيانات من أجل فهم مقدار ونوع المعلومات الموجودة بالفعل وما هي المعلومات التي يمكن استردادها من مصادر خارجية.

التصميم. خلال مرحلة التصميم ، السؤال الرئيسي هو: كيف ستعمل DSS؟ وبالتالي ، يتم تحديد بنية النظام بالكامل ، من خلال تحديد منصات تكنولوجيا الأجهزة ، وهيكل الشبكة ، وأدوات البرمجيات لتطوير التطبيقات وقاعدة البيانات المحددة التي سيتم استخدامها. من الضروري أيضًا تحديد التفاعلات مع المستخدمين بالتفصيل ، عن طريق أقنعة الإدخال ، والتصورات الرسومية على الشاشة والتقارير المطبوعة. أصبح متصفح الويب في السنوات الأخيرة أداة تفاعل مهمة ، وساهم بالتأكيد في تنسيق وتبسيط المشكلات المتعلقة بالاتصال بين العاملين في مجال المعرفة وأجهزة الكمبيوتر. هناك جانب آخر يجب توضيحه خلال مرحلة التصميم وهو خيار الصنع أو الشراء - سواء كان التعاقد من الباطن لتنفيذ DSS إلى أطراف ثالثة ، كليًا أو جزئيًا.

التطبيق. بمجرد وضع المواصفات ، حان وقت التنفيذ والاختبار والتثبيت الفعلي ، عندما يتم طرح DSS وتشغيله. يمكن إرجاع أي مشاكل واجهتها في هذه المرحلة الأخيرة إلى أساليب إدارة المشروع. يتعلق جانب آخر من مرحلة التنفيذ ، والذي غالبًا ما يتم تجاهله ، بالتأثير العام على المنظمة الذي يحدده النظام الجديد. يجب مراقبة هذه التأثيرات باستخدام تقنيات إدارة التغيير ، والتأكد من عدم شعور أي شخص بالاستبعاد من عملية الابتكار التنظيمي ورفض DSS.

النتائج. في بعض الأحيان ، قد لا يصل المشروع إلى نتيجة ناجحة ، أو قد لا ينجح في تحقيق التوقعات ، أو قد يتحول إلى فشل كامل.

ومع ذلك ، هناك طرق لتقليل مخاطر الفشل. يعتمد أهمها على استخدام التطوير السريع للنماذج الأولية حيث ، بدلاً من تنفيذ النظام ككل ، يتمثل النهج في تحديد سلسلة من الأنظمة الفرعية المستقلة ذات القدرات المحدودة ، وتطوير هذه الأنظمة الفرعية خطوة بخطوة حتى النهاية تم الوصول إلى المرحلة المقابلة لـ DSS المطورة بالكامل.

يوفر التطوير السريع للنماذج الأولية مزايا واضحة. يمكن في الواقع تطوير كل نظام فرعي بسرعة أكبر وبالتالي فهو متاح بسهولة أكبر.

علاوة على ذلك ، عندما يتم إصدار نظام فرعي للمستخدمين ، فمن الممكن التحقق من توافقه مع الغرض المقصود واختبار وظائفه حتى لو لم يتم تطويرها بالكامل. ومن ثم ، فإن التطور التطوري لـ DSS يسمح بتقليل مخاطر الفشل. علاوة على ذلك ، تتيح الاختبارات الوسيطة إمكانية تصحيح معظم أخطاء التصميم على الفور. في حالة وجود تناقض واضح بين ميزات النموذج الأولي الذي تم تطويره وتوقعات المستخدمين ، فمن الممكن حتى إيقاف المشروع مبكرًا.

يمكن اعتماد طرق تطوير أخرى بشكل فعال من أجل تسريع تنفيذ البرنامج. وتشمل هذه تقنيات التطوير السريع وتقنيات البرمجة المتطرفة.

هناك جانب آخر لا ينبغي إغفاله وهو الإدارة الدورية ومراجعة نظام دعم القرار. نظراً لأنه لا ينبغي اعتبار أي تطبيق برمجي منتهياً تماماً وغير قابل للتغيير بمرور الوقت ، فإن هذا الأمر أكثر أهمية بالنسبة لأداة تهدف إلى دعم العاملين في مجال المعرفة في عملية صنع القرار في المواقف الديناميكية المتغيرة بسرعة. نتيجة لذلك ، من الضروري تطوير DSS من خلال توفير التغييرات والتعديلات المستقبلية.

في هذا الصدد ، من الممكن تحديد ميزتين واضحتين توفرهما طرق التطوير الموصوفة حتى الآن. أولاً ، إذا تم تطوير DSS باستخدام تقنيات النماذج الأولية السريعة ، فمن المحتمل أن تكون أكثر قابلية للتمديدات والمراجعات الدورية. علاوة على ذلك ، فإن النماذج الرياضية المستخدمة لتطوير DSS مناسبة لدمج التغييرات والتحديثات المتطورة التي غالبًا ما تقلل إلى تعديلات بسيطة في التعبيرات التي تصف النموذج.

نختتم هذا القسم من خلال النظر في العوامل الحاسمة الرئيسية التي قد تؤثر على درجة نجاح DSS.

اندماج. يتطلب تصميم وتطوير DSS عددًا كبيرًا من المنهجيات والأدوات والنماذج والأفراد والعمليات التنظيمية للعمل في وئام. ينتج عن هذا مشروع معقد للغاية يتطلب كفاءات متنوعة. وبالتالي فإن دور مكامل النظام ضروري. مكامل النظام هو خبير في جميع الجوانب التي ينطوي عليها تطوير DSS ، مثل معماريات نظام المعلومات ، وعمليات صنع القرار ، والنماذج الرياضية وطرق الحل. عادة ما يؤدي هذا الدور طرف ثالث قد يكون له أيضًا تأثير إيجابي كعامل ابتكار قادر على التغلب على معظم مقاومة التغيير التي تنشأ غالبًا في كل مؤسسة.

المشاركة. يعد الاستبعاد أو التهميش من فريق المشروع من العاملين في مجال المعرفة الذين سيستخدمون النظام بالفعل بمجرد تنفيذه خطأ يحدث أحيانًا أثناء تصميم وتطوير DSS.
في كثير من الحالات ، يحدث هذا بسبب اعتبار DSS عن طريق الخطأ مجرد تطبيق كمبيوتر ، يتم تخصيص تطويره فقط أو بشكل أساسي لقسم نظم المعلومات. وعلى العكس من ذلك ، فإن إشراك صانعي القرار والمستخدمين أثناء عملية التطوير له أهمية قصوى لتجنب ميلهم إلى رفض أداة يرون أنها غريبة. من الضروري أيضًا تعزيز عمليات الاتصال غير الرسمية ، خاصة أثناء مرحلتي التصميم والتنفيذ للنماذج الأولية.
عدم التأكد (الشك) بشكل عام ، التكاليف ليست مصدر قلق كبير في تنفيذ DSS ، وميزة استنباط قرارات أكثر فاعلية تعوض إلى حد كبير تكاليف التطوير المتكبدة. بالطبع ، من المناسب تقليل عدم اليقين في المشروع من خلال النماذج الأولية ، وسهولة الاستخدام ، واختبارات النظام خلال المراحل الأولية والتنفيذ التطوري.

2.6 ملاحظات وقراءات

من بين المجلدات العديدة المخصصة لأنظمة دعم القرار ، نود اقتراح القراءات التالية: Malach (1995) Klein and Methlie ، Turban et al. (2005) ، ((2000)) و Dhar and Stein (1997). لمراجعة حديثة ، انظر Moss and Atre (2003) .Shim et al. (2002) يحللان عملية تطوير DSS. من بين المساهمات الرئيسية التي تعود إلى أسس DSS ، انظر Simon (1969، 1977)، Gorry and Scott Morton (1971)، Keen and Scott Morton (1978) and Sprague (1980).

الفصل الثالث: مستودع البيانات

Data Warehouse

كما لوحظ في الفصل 2 ، منذ منتصف التسعينيات ، كان هناك شعور بالحاجة إلى الفصل المنطقي والمادي بين قواعد البيانات التي تغذي بيانات الإدخال في أنظمة دعم القرار وبنى استخبارات الأعمال من ناحية ، وأنظمة المعلومات التشغيلية من ناحية أخرى.

سنصف في هذا الفصل ميزات مستودعات البيانات ومجموعات البيانات ، مع توضيح العوامل التي أدت إلى تصورها ، وإبراز الاختلافات الرئيسية بينها وبين أنظمة التشغيل ، ومناقشة المتطلبات المتعلقة بجودة البيانات. ثم سنقوم بفحص بنية مستودع البيانات ، مع الإشارة إلى دور أدوات ETL والبيانات الوصفية. سيخصص الجزء الأخير من الفصل لعمليات المعالجة التحليلية عبر الإنترنت والتحليلات التي يمكن إجراؤها باستخدام مكعبات متعددة الأبعاد وتسلسلات هرمية للمفاهيم.

سنركز مناقشتنا على أهداف ووظائف مستودع البيانات ، ونتجنب عمداً المشكلات الفنية المتعلقة بتطويرها. بالنسبة لهذه الأخيرة ، قد نحيل القراء إلى كتب أكثر تحديدًا مخصصة بالكامل للموضوع ، المشار إليه في القسم الأخير من الفصل.

3.1 تعريف مستودع البيانات Definition of data warehouse

كما يوحي اسمه ، مستودع البيانات هو المستودع الأول للبيانات المتاحة لتطوير بنى ذكاء الأعمال وأنظمة دعم القرار. يشير مصطلح تخزين البيانات إلى المجموعة الكاملة من الأنشطة المترابطة المشاركة في تصميم وتنفيذ واستخدام مستودع البيانات.

من الممكن تحديد ثلاث فئات رئيسية لتغذية البيانات في مستودع البيانات: البيانات الداخلية والبيانات الخارجية والبيانات الشخصية.

1. **البيانات الداخلية.** يتم تخزين البيانات الداخلية في معظمها في قواعد البيانات ، والتي يشار إليها باسم أنظمة المعاملات أو أنظمة التشغيل ، والتي تعد العمود الفقري لنظام معلومات المؤسسة. يتم جمع البيانات الداخلية من خلال تطبيقات المعاملات التي تترأس بشكل روتيني عمليات الشركة ، مثل الإدارة والمحاسبة والإنتاج والخدمات اللوجستية.

تسمى هذه المجموعة من تطبيقات برامج المعاملات بتخطيط موارد المؤسسة (enterprise resource planning, ERP).

عادةً ما تتعامل البيانات المخزنة في أنظمة التشغيل مع الكيانات الرئيسية المشاركة في عمليات الشركة ، وهي العملاء والمنتجات والمبيعات والموظفون والموردون. تأتي هذه البيانات عادة من مكونات مختلفة لنظام المعلومات:

- أنظمة المكتب الخلفي ، التي تجمع سجلات المعاملات الأساسية مثل الطلبات والفواتير والمخزونات وبيانات الإنتاج واللوجستيات.
- أنظمة المكاتب الأمامية ، التي تحتوي على البيانات الناشئة عن أنشطة مركز الاتصال ، ومساعدة العملاء ، وتنفيذ الحملات التسويقية ؛
- الأنظمة المستندة إلى الويب ، والتي تجمع معاملات البيع على مواقع التجارة الإلكترونية على شبكة الإنترنت ، وزيارات مواقع الويب ، والبيانات المتاحة على النماذج المملوءة من قبل العملاء الحاليين والمحتملين.

2. **بيانات خارجية.** هناك العديد من مصادر البيانات الخارجية التي يمكن استخدامها لتوسيع ثروة المعلومات المخزنة في قواعد البيانات الداخلية. على سبيل المثال ، تقوم بعض الوكالات بجمع وإتاحة البيانات المتعلقة بالمبيعات وحصة السوق وتنبؤات الاتجاهات المستقبلية لصناعات أعمال معينة ، فضلاً عن المؤشرات الاقتصادية والمالية. توفر الوكالات الأخرى استبيانات سوق البيانات وآراء المستهلكين التي تم جمعها من خلال الاستبيانات.

يتم توفير مصدر مهم آخر للبيانات الخارجية من خلال أنظمة المعلومات الجغرافية (GIS) ، والتي تمثل مجموعة من التطبيقات للحصول على البيانات الإقليمية وتنظيمها وتخزينها وتقديمها. تحتوي على معلومات تتعلق بالكيانات التي لها موقع جغرافي محدد. لذلك يرتبط كل كيان بإحداثيات خطوط الطول والعرض ، إلى جانب بعض السمات الأخرى ، التي تنشأ عادةً من قواعد البيانات العلائقية وتعتمد فعلياً على مجال التطبيق. وبالتالي ، تسمح هذه البيانات بإجراء تحليلات خاصة بالموضوع على البيانات المرتبطة بالعناصر الجغرافية والنتائج المراد تصورها بيانياً.

3. **بيانات شخصية.** في معظم الحالات ، يعتمد صانعو القرار الذين يقومون بتحليل ذكاء الأعمال أيضًا على المعلومات والتقييمات الشخصية المخزنة داخل أوراق العمل أو قواعد البيانات المحلية الموجودة في أجهزة الكمبيوتر الخاصة بهم. يعد استرجاع هذه المعلومات وتكاملها مع البيانات المنظمة من المصادر الداخلية والخارجية أحد أهداف أنظمة إدارة المعرفة.

يشار إلى التطبيقات البرمجية الموجودة في قلب أنظمة التشغيل بمعالجة المعاملات عبر الإنترنت (on-line transaction processing, OLTP). من ناحية أخرى ، فإن المجموعة الكاملة من الأدوات التي تهدف إلى إجراء تحليلات ذكاء الأعمال ودعم عمليات اتخاذ القرار تأتي باسم المعالجة التحليلية عبر الإنترنت (on-line analytical processing, OLAP). لذلك يمكننا أن نفترض أن وظيفة مستودع البيانات هي توفير بيانات الإدخال لتطبيقات OLAP.

هناك عدة أسباب لتنفيذ مستودع بيانات بشكل منفصل عن قواعد البيانات التي تدعم تطبيقات OLTP في المؤسسة. من بينها ، نذكر هنا الأكثر صلة بالموضوع.

- **اندماج Integration**. في كثير من الحالات ، يجب أن تصل أنظمة دعم القرار إلى المعلومات الناشئة من عدة مصادر بيانات ، موزعة عبر أجزاء مختلفة من المؤسسة أو مشتقة من مصادر خارجية. ومن ثم ، يلزم وجود مستودع بيانات يدمج مصادر متعددة وغير متجانسة في كثير من الأحيان لتعزيز وتسهيل الوصول إلى المعلومات. يمكن تحقيق تكامل البيانات عن طريق تقنيات مختلفة - على سبيل المثال ، باستخدام طرق تشفير موحدة ، والتحويل إلى وحدات قياس قياسية وتحقيق التجانس الدلالي للمعلومات.
- **جودة Quality.** يتم فحص البيانات المنقولة من أنظمة التشغيل إلى مستودع البيانات وتصحيحها من أجل الحصول على معلومات موثوقة وخالية من الأخطاء ، قدر الإمكان. وغني عن القول ، أن هذا يزيد من القيمة العملية لأنظمة ذكاء الأعمال المطورة بدءًا من البيانات الموجودة في مستودع البيانات.
- **كفاءة Efficiency.** قد تكون الاستفسارات/الأستعلامات Queries التي تهدف إلى استخراج المعلومات لتحليل ذكاء الأعمال مرهقة من حيث موارد الحوسبة ووقت المعالجة. ونتيجة لذلك ، إذا تم توجيه استعلام "قاتل" إلى أنظمة المعاملات ، فمن شأنه أن يضر بشدة بالكفاءة التي تتطلبها تطبيقات تخطيط موارد المؤسسة ، مع عواقب سلبية

على الأنشطة الروتينية للشركة. الحل الأفضل هو توجيه الاستعلامات المعقدة لتحليلات OLAP إلى مستودع البيانات ، منفصلة فعليًا عن أنظمة التشغيل.

- **التمددية Extendability.** تمتد البيانات المخزنة في أنظمة المعاملات على مدى فترة زمنية محدودة في الماضي. في الواقع ، نظرًا للقيود المفروضة على سعة الذاكرة ، تتم إزالة البيانات المتعلقة بالفترات الماضية بانتظام من أنظمة OLTP ويتم أرشفتها بشكل دائم في أجهزة التخزين الكبيرة غير المتصلة بالإنترنت ، مثل أقراص DVD أو الأشرطة المغناطيسية. من ناحية أخرى ، تحتاج أنظمة ذكاء الأعمال ونماذج التنبؤ إلى الوصول إلى جميع البيانات السابقة المتاحة حتى تتمكن من فهم الاتجاهات واكتشاف الأنماط المتكررة. هذا ممكن بسبب قدرة مخازن البيانات على الاحتفاظ بالمعلومات التاريخية.

في ضوء الملاحظات السابقة ، يمكننا تعريف مستودع البيانات على أنه مجموعة من البيانات التي تدعم عمليات صنع القرار وأنظمة ذكاء الأعمال التي لها الخصائص التالية.

1. **كياني المنحى Entity Oriented.** تتعلق البيانات الموجودة في مستودع البيانات بشكل أساسي بالكيانات الرئيسية ذات الأهمية للتحليل ، مثل المنتجات والعملاء والطلبات والمبيعات. من ناحية أخرى ، تكون أنظمة المعاملات أكثر توجهاً نحو الأنشطة التشغيلية وتستند إلى كل معاملة فردية مسجلة بواسطة تطبيقات تخطيط موارد المؤسسة. أثناء تحليل ذكاء الأعمال ، يتيح التوجه نحو الكيانات تقييم أداء الشركة بسهولة أكبر واكتشاف أي مصدر محتمل لعدم الكفاءة.
2. **المدمجة Integrated.** يتم دمج البيانات الناشئة من المصادر المختلفة وتجانسها حيث يتم تحميلها في مستودع البيانات. على سبيل المثال ، يتم تنسيق وحدات القياس والتشفير وجعلها متسقة.
3. **تغير الوقت Time Variant.** يتم تصنيف جميع البيانات التي يتم إدخالها في مستودع البيانات بالفترة الزمنية التي تشير إليها. يمكننا ربط البيانات المخزنة في مستودع البيانات إلى حد ما بسلسلة من صور اللقطة غير المتطايرة ، التي تم التقاطها في أوقات متتالية وتحمل ملصق الفترة المرجعية. نتيجة لذلك ، يعد البعد الزمني في أي مستودع بيانات عنصرًا مهمًا يلعب دورًا رئيسيًا. وبهذه الطريقة ، قد تقوم تطبيقات دعم القرار بتطوير تحليلًا تاريخيًا لهذه التوجهات.

4. **البقاء Persistent.** بمجرد تحميلها في مستودع بيانات ، لا يتم عادةً تعديل البيانات أكثر ويتم الاحتفاظ بها بشكل دائم. تعمل هذه الميزة على تسهيل تنظيم الوصول للقراءة فقط من قبل المستخدمين وتبسيط عملية التحديث ، وتجنب التزامن الذي يمثل أهمية بالغة لأنظمة التشغيل.
5. **موحّد Consolidated.** عادةً ما يتم الحصول على بعض البيانات المخزنة في مستودع البيانات كملخصات جزئية للبيانات الأولية التي تنتمي إلى أنظمة التشغيل التي تنشأ منها. على سبيل المثال ، قد تقوم شركة الهاتف المحمول بتخزين التكلفة الإجمالية للمكالمات التي أجراها كل عميل في مستودع بيانات في أسبوع ، مقسمة حسب طرق المرور ونوع الخدمة المحددة ، بدلاً من تخزين المكالمات الفردية المسجلة بواسطة أنظمة التشغيل. سبب هذا الدمج ذو شقين: من ناحية ، تقليل المساحة المطلوبة لتخزين البيانات المتراكمة على مر السنين في مستودع البيانات ؛ من ناحية أخرى ، قد تكون المعلومات الموحدة قادرة على تلبية احتياجات أنظمة ذكاء الأعمال بشكل أفضل.
6. **غير طبيعي- معدلة Denormalized.** على عكس قواعد البيانات التشغيلية ، لا يتم تنظيم البيانات المخزنة في مستودع البيانات بالشكل العادي ولكن يمكن بدلاً من ذلك توفير التكرار ، للسماح بوقت استجابة أقصر للاستعلامات المعقدة.

يمثل التقسيم Granularity أعلى مستوى من التفاصيل يتم التعبير عنه بواسطة البيانات الأولية الموجودة في مستودع البيانات ، والتي يشار إليها أيضًا بالبيانات الذرية atomic data. من الواضح أن دقة مستودع البيانات لا يمكن أن تتجاوز تلك الخاصة بمصادر البيانات الأصلية. بشكل عام ، هو أقل بشكل صارم بسبب الدمج الذي يهدف إلى تقليل شغل التخزين ، كما هو موضح أعلاه.

تختلف فلسفة التصميم الكامنة وراء مستودعات البيانات تمامًا عن تلك المعتمدة لقواعد البيانات التشغيلية. يلخص الجدول 3.1 الاختلافات الرئيسية بين أنظمة OLTP و OLAP.

الجدول 3.1 الاختلافات بين خصائص أنظمة OLTP و OLAP

Characteristic	OLTP	OLAP
volatility	dynamic data	static data
timeliness	current data only	current and historical data
time dimension	implicit and current	explicit and variant
granularity	detailed data	aggregated and consolidated data
updating	continuous and irregular	periodic and regular
activities	repetitive	unpredictable
flexibility	low	high
performance	high, few seconds per query	may be low for complex queries
users	employees	knowledge workers
functions	operational	analytical
purpose of use	transactions	complex queries and decision support
priority	high performance	high flexibility
metrics	transaction rate	effective response
size	megabytes to gigabytes	gigabytes to terabytes

3.1.1 سوق/متجر البيانات Data Marts

تعد متاجر البيانات Data Marts أنظمة تجمع جميع البيانات المطلوبة من قبل قسم معين في الشركة ، مثل التسويق أو الخدمات اللوجستية ، لغرض إجراء تحليلات ذكاء الأعمال وتنفيذ تطبيقات دعم القرار الخاصة بالوظيفة نفسها. لذلك ، يمكن اعتبار سوق البيانات بمثابة مستودع بيانات وظيفي أو إداري بحجم أصغر ونوع أكثر تحديدًا من مستودع بيانات الشركة الإجمالي.

لذلك ، يحتوي متجر البيانات على مجموعة فرعية من البيانات المخزنة في مستودع بيانات الشركة ، والتي عادةً ما تتكامل مع البيانات الأخرى التي يمتلكها قسم الشركة المسؤول عن سوق البيانات ويرى أنها مهمة. على سبيل المثال ، سوف يحتوي متجر بيانات التسويق على البيانات المستخرجة من مستودع البيانات المركزي ، مثل معلومات العملاء ومعاملات المبيعات ، ولكن أيضًا البيانات الإضافية المتعلقة بوظيفة التسويق ، مثل نتائج الحملات التسويقية التي تم تشغيلها في الماضي.

وبالتالي ، فإن مستودعات البيانات ومتاجر البيانات تشتركان في نفس الإطار التكنولوجي. من أجل تنفيذ تطبيقات ذكاء الأعمال ، تفضل بعض الشركات أن تصمم وتطور بطريقة تدريجية سلسلة من متاجر البيانات المتكاملة بدلاً من مستودع بيانات مركزي ، من أجل تقليل وقت التنفيذ والشكوك المرتبطة بالمشروع.

3.1.2 جودة البيانات Data quality

تعد الحاجة إلى التحقق من جودة البيانات والحفاظ عليها وتحسينها مصدر قلق دائم للمسؤولين عن تصميم وتحديث مستودع البيانات.
المشاكل الرئيسية التي قد تعرض للخطر صحة وسلامة البيانات موضحة في الجدول 3.2.

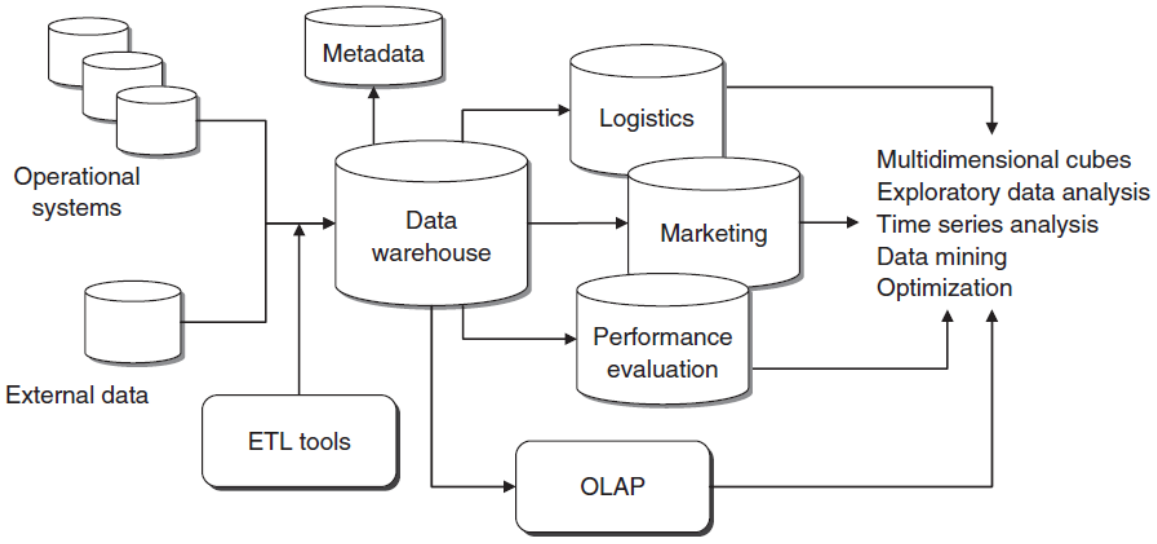
الجدول 3.2 سلامة البيانات: المشاكل والأسباب والعلاجات

Problem	Cause	Remedy
incorrect data	data collected without due care	systematic checking of input data
	data entered incorrectly	data entry automation
	uncontrolled modification of data	implementation of a safety program for access and modifications
data not updated	data collection does not match user needs	timely updating and collection of data
		retrieval of updated data from the web
missing data	failure to collect the required data	identification of data needed via preliminary analysis and estimation of missing data

بشكل عام ، يمكننا تحديد العوامل الرئيسية التالية التي قد تؤثر على جودة البيانات.

- **الدقة Accuracy.** لكي تكون مفيدة في التحليلات اللاحقة ، يجب أن تكون البيانات عالية الدقة. على سبيل المثال ، من الضروري التحقق من تمثيل الأسماء والتشفيرات بشكل صحيح وأن القيم ضمن النطاقات المسموح بها.
- **الاكتمال Completeness.** من أجل تجنب المساومة على دقة تحليلات ذكاء الأعمال ، يجب ألا تتضمن البيانات عددًا كبيرًا من القيم المفقودة. ومع ذلك ، يجب على المرء أن يضع في اعتباره أن معظم تقنيات التعلم واستخراج البيانات قادرة على التقليل بطريقة قوية من آثار عدم الاكتمال الجزئي في البيانات.
- **التناسق Consistency.** يجب أن يكون شكل ومحتوى البيانات متسقًا عبر مصادر البيانات المختلفة بعد إجراءات التكامل ، فيما يتعلق بوحدات العملة والقياس.

- **حسن التوقيت Timeliness.** يجب تحديث البيانات بشكل متكرر ، بناءً على أهداف التحليل. من المعتاد ترتيب تحديث لمستودع البيانات بانتظام على أساس يومي أو أسبوعي على الأكثر.
- **عدم التكرار Non-Redundancy.** يجب تجنب تكرار البيانات والتكرار من أجل منع ضياع الذاكرة والتناقضات المحتملة. ومع ذلك ، يمكن نسخ البيانات عندما يؤدي إلغاء تسوية مستودع البيانات إلى تقليل أوقات الاستجابة للاستعلامات المعقدة.
- **ملاءمة Relevance.** يجب أن تكون البيانات ذات صلة باحتياجات نظام ذكاء الأعمال من أجل إضافة قيمة حقيقية للتحليلات التي سيتم إجراؤها لاحقًا.
- **أمكانية التفسيرInterpretability.** يجب أن يكون معنى البيانات مفهوماً جيداً ومفسراً بشكل صحيح من قبل المحللين ، بناءً على الوثائق المتوفرة في البيانات الوصفية التي تصف مستودع البيانات ، كما هو موضح في القسم 3.2.2.
- **إمكانية الوصول Accessibility.** يجب أن تكون البيانات سهلة الوصول للمحللين وتطبيقات دعم القرار.



الشكل 3.1 بنية ووظائف مستودع البيانات

3.2 بنية مستودع البيانات Data warehouse architecture

تتضمن البنية المرجعية لمستودع البيانات ، الموضحة في الشكل 3.1 ، المكونات الوظيفية الرئيسية التالية.

1. مستودع البيانات نفسه ، جنبًا إلى جنب مع مجموعات البيانات الإضافية ، التي تحتوي على البيانات والوظائف التي تسمح بالوصول إلى البيانات وتصورها وربما تعديلها.
2. تطبيقات الحصول على البيانات Data acquisition applications ، والمعروفة أيضًا باسم الاستخراج والتحويل والتحميل (ETL) أو الأدوات الخلفية back-end tools، والتي تسمح باستخراج البيانات وتحويلها وتحميلها في مستودع البيانات.
3. تطبيقات ذكاء الأعمال ودعم القرار ، والتي تمثل الواجهة الأمامية front-end وتسمح للعاملين في مجال المعرفة بإجراء التحليلات وتصور النتائج.

ينطبق التمييز ثلاثي المستويات على البنية الموضحة في الشكل 3.1 حتى من منظور تكنولوجي.

1. مستوى مصادر البيانات وأدوات ETL ذات الصلة والتي يتم تثبيتها عادةً على خادم واحد أو أكثر.
2. مستوى مستودع البيانات وأي سوق بيانات ، من المحتمل أن يكون متاحًا على خادم واحد أو أكثر أيضًا ، ومنفصل عن تلك التي تحتوي على مصادر البيانات. يتضمن هذا المستوى الثاني أيضًا البيانات الوصفية التي توثق أصل ومعنى السجلات المخزنة في مستودع البيانات.
3. مستوى التحليلات التي تزيد من قيمة المعلومات الموجودة في مستودع البيانات من خلال الاستعلام وإعداد التقارير وربما أدوات دعم القرار المعقدة. عادة ما يتم العثور على تطبيقات ارتباط إنتل الأعمال وتحليل دعم القرار على خوادم منفصلة أو مباشرة على كمبيوتر العميل المستخدم من قبل المحللين والعاملين في مجال المعرفة.

يتم أيضًا اعتماد أنظمة إدارة قواعد البيانات database management system نفسها المستخدمة لتطوير أنظمة المعاملات لتنفيذ مستودعات البيانات ومتاجر البيانات. نظرًا لمتطلبات

الاستجابة التي أثارتها الاستعلامات المعقدة الموجهة إلى مستودع البيانات ، فإن أنظمة إدارة قواعد البيانات المستخدمة لتخزين البيانات تخضع لهيكلة ومعايير مختلفة فيما يتعلق بأنظمة المعاملات.

يمكن تنفيذ مستودع البيانات وفقًا لأساليب التصميم المختلفة: من أعلى إلى أسفل top-down ومن أسفل إلى أعلى bottom-up ومختلط mixed..

1. **من أعلى إلى أسفل top-down.** تعتمد المنهجية التنازلية على التصميم العام لمستودع البيانات ، وبالتالي فهي أكثر منهجية. ومع ذلك ، فإنه ينطوي على أوقات تطوير أطول ومخاطر أعلى لعدم اكتمالها ضمن الجدول الزمني نظرًا لأن مستودع البيانات بأكمله قيد التطوير بالفعل.
2. **تصاعدي bottom-up.** تعتمد الطريقة التصاعدية على استخدام النماذج الأولية ، وبالتالي يتم إجراء امتدادات النظام وفقًا لمخطط تدريجي. عادة ما يكون هذا النهج أسرع ، ويوفر نتائج ملموسة أكثر ولكنه يفتقر إلى رؤية شاملة للنظام بأكمله ليتم تطويره.
3. **مختلطMixed.** تعتمد المنهجية المختلطة على التصميم العام لمخزن البيانات ، ولكنها تتابع بعد ذلك نهج النماذج الأولية ، من خلال التنفيذ المتسلسل لأجزاء مختلفة من النظام بأكمله. هذا النهج عملي للغاية ويفضل عادة ، لأنه يسمح باتخاذ خطوات صغيرة وخاضعة للرقابة مع الأخذ في الاعتبار الصورة الكاملة.

يمكن تلخيص خطوات تطوير مستودع البيانات أو سوق البيانات على النحو التالي.

1. يتم تحديد عملية واحدة أو أكثر داخل المنظمة ليتم تمثيلها في مستودع البيانات ، مثل المبيعات أو الخدمات اللوجستية أو المحاسبة.
2. يتم تحديد الدقة المناسبة لتمثيل العمليات المختارة والمستوى الذري للبيانات.
3. يتم بعد ذلك اختيار التدابير ذات الصلة التي سيتم التعبير عنها في جداول الحقائق للتحليل متعدد الأبعاد ، كما هو موضح في القسم 3.3.
4. أخيرًا ، يتم تحديد أبعاد جداول الحقائق.

3.2.1 أدوات ETL

يشير ETL إلى أدوات البرامج المخصصة لأداء ثلاث وظائف رئيسية بطريقة تلقائية: استخراج البيانات وتحويلها وتحميلها في مستودع البيانات.

- **استخلاص Extraction..** خلال المرحلة الأولى ، يتم استخراج البيانات من المصادر الداخلية والخارجية المتاحة. يمكن إجراء تمييز منطقي بين الاستخراج الأولي ، حيث يتم إدخال البيانات المتاحة المتعلقة بجميع الفترات السابقة في مستودع البيانات الفارغ ، وعمليات الاستخراج الإضافية اللاحقة التي تعمل على تحديث مستودع البيانات باستخدام البيانات الجديدة التي أصبحت متاحة بمرور الوقت. يعتمد اختيار البيانات التي سيتم استيرادها على تصميم مستودع البيانات ، والذي يعتمد بدوره على المعلومات التي تحتاجها تحليلات ذكاء الأعمال وأنظمة دعم القرار التي تعمل في مجال تطبيق معين.
- **تحويل Transformation.** الهدف من مرحلة التنظيف والتحويل هو تحسين جودة البيانات المستخرجة من المصادر المختلفة ، من خلال تصحيح التناقضات وعدم الدقة والقيم المفقودة. بعض أوجه القصور الرئيسية التي تمت إزالتها أثناء مرحلة تنقية البيانات هي:
  - عدم الاتساق بين القيم المسجلة في سمات مختلفة لها نفس المعنى.
  - ازدواجية البيانات.
  - بيانات مفقودة
  - وجود قيم غير مقبولة.

أثناء مرحلة التنظيف ، يتم تطبيق القواعد التلقائية المحددة مسبقًا لتصحيح معظم الأخطاء المتكررة. في كثير من الحالات ، تُستخدم القواميس ذات المصطلحات الصالحة لاستبدال المصطلحات التي يُفترض أنها غير صحيحة ، بناءً على مستوى التشابه. علاوة على ذلك ، أثناء مرحلة التحويل ، تحدث تحويلات بيانات إضافية من أجل ضمان التجانس والتكامل فيما يتعلق بمصادر البيانات المختلفة. علاوة على ذلك ، يتم إجراء تجميع البيانات ودمجها من أجل الحصول على الملخصات التي ستقلل من وقت الاستجابة المطلوب من خلال الاستعلامات والتحليلات اللاحقة التي تم تخصيص مستودع البيانات لها.

- **تحميل Loading.** أخيرًا ، بعد استخراج البيانات وتحويلها ، يتم تحميل البيانات في جداول مستودع البيانات لإتاحتها للمحللين وتطبيقات دعم القرار.

3.2.2 البيانات الوصفية Metadata

من أجل توثيق documentation معنى البيانات الواردة في مستودع البيانات ، يوصى بإعداد هيكل معلومات محدد ، يُعرف باسم البيانات الوصفية ، أي البيانات التي تصف البيانات. تشير البيانات الوصفية لكل سمة من سمات مستودع البيانات إلى المصدر الأصلي للبيانات ومعناها والتحولات التي خضعت لها. يجب أن يتم تحديث الوثائق المقدمة بواسطة البيانات الوصفية باستمرار ، من أجل عكس أي تعديل في هيكل مستودع البيانات. يجب أن يكون التوثيق متاحًا بشكل مباشر لمستخدمي مستودع البيانات ، من الناحية المثالية من خلال متصفح الويب ، وفقًا لحقوق الوصول المتعلقة بأدوار كل محلل.

على وجه الخصوص ، يجب أن تؤدي البيانات الوصفية المهام الإعلامية التالية:

- توثيق هيكل مستودع البيانات: التخطيط ، وجهات النظر المنطقية ، والأبعاد ، والتسلسل الهرمي ، والبيانات المشتقة ، وتوطين أي سوق للبيانات ؛
- توثيق لأصول البيانات التي تم الحصول عليها عن طريق وضع علامات على مصادر البيانات التي تم استخلاص البيانات منها وعن طريق وصف أي تشكيل متحول تم إجراؤه على البيانات نفسها ؛
- قائمة تحتفظ بإحصائيات الاستخدام لمستودع البيانات ، من خلال الإشارة إلى عدد مرات الوصول إلى الحقل أو العرض المنطقي التي تم إجراؤها ؛
- توثيق المعنى العام لمستودع البيانات فيما يتعلق بمجال التطبيق ، من خلال توفير تعريف المصطلحات المستخدمة ، ووصف كامل لخصائص البيانات وملكية البيانات وسياسات التحميل.

3.3 المكعبات والتحليل متعدد الأبعاد

Cubes and multidimensional analysis

يعتمد تصميم مستودعات البيانات ومتاجر البيانات على نموذج متعدد الأبعاد لتمثيل البيانات والذي يوفر على الأقل ميزتين رئيسيتين: على الجانب الوظيفي ، يمكن أن يضمن أوقات الاستجابة السريعة حتى للاستعلامات المعقدة ، بينما من الجانب المنطقي تتطابق الأبعاد بشكل طبيعي مع المعايير التي يتبعها عمال المعرفة لإجراء تحليلاتهم.

يعتمد التمثيل متعدد الأبعاد على مخطط نجمة يحتوي على نوعين من جداول البيانات: جداول الأبعاد dimension tables وجداول الحقائق fact tables.

جداول الأبعاد dimension tables. بشكل عام ، ترتبط الأبعاد بالكيانات التي تدور حولها عمليات المؤسسة. ثم تتوافق جداول الأبعاد مع الكيانات الأساسية الموجودة في مستودع البيانات ، وفي معظم الحالات يتم اشتقاقها مباشرةً من الجداول الرئيسية master tables المخزنة في أنظمة OLTP ، مثل العملاء والمنتجات والمبيعات والمواقع والوقت. غالبًا ما يتم تنظيم كل جدول أبعاد داخليًا وفقًا للعلاقات الهرمية hierarchical relationships. على سبيل المثال ، عادةً ما يستند البعد الزمني إلى تسلسلين هرمين رئيسيين: {يوم ، أسبوع ، سنة} و {يوم ، شهر ، ربع سنة ، سنة}. وبالمثل ، قد يتم تنظيم بُعد الموقع بشكل هرمي كـ {شارع ، رمز بريدي ، مدينة ، مقاطعة ، منطقة ، بلد ، منطقة}. المنتجات بدورها لها هياكل هرمية مثل {type، family، item} في الصناعة التحويلية و {department، category، item} في صناعة البيع بالتجزئة. بطريقة ما ، تحدد الأبعاد مسبقًا المسارات الرئيسية التي من المفترض أن يتم تطوير تحليلات OLAP على أساسها.

جداول الحقائق Fact tables. تشير جداول الحقائق عادةً إلى المعاملات وتحتوي على نوعين من البيانات:

- روابط لجداول الأبعاد المطلوبة للإشارة بشكل صحيح إلى المعلومات الواردة في كل جدول حقائق.
- القيم الرقمية للسمات التي تميز المعاملات المقابلة والتي تمثل الهدف الفعلي لتحليلات OLAP اللاحقة.

على سبيل المثال ، قد يحتوي جدول الحقائق على معاملات المبيعات والإشارة إلى جداول أبعاد متعددة ، مثل العملاء ونقاط البيع والمنتجات والموردين والوقت. مقاييس الفائدة المقابلة هي سمات مثل كمية العناصر المباعة وسعر الوحدة والخصم. في هذا المثال ، يسمح جدول الحقائق للمحللين بتقييم اتجاهات المبيعات بمرور الوقت ، إما إجمالاً ، أو إحالة إلى عميل واحد ، أو الإشارة إلى مجموعة من العملاء ، والتي يمكن تحديدها من خلال أي تسلسل هرمي ناتج عن جدول الأبعاد المرتبط بـ عملاء.

قد يقوم المحلل أيضًا بتقييم الاتجاه بمرور الوقت لنسب المبيعات بالنسبة للعملاء الموجودين في منطقة معينة.

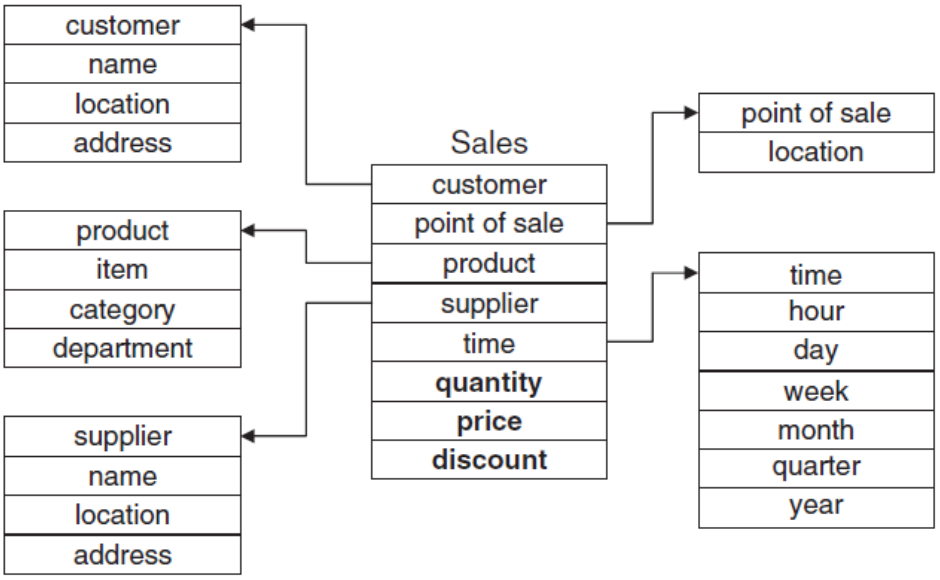
يوضح الشكل 3.2 مخطط النجوم المرتبط بجدول الحقائق الذي يمثل معاملات المبيعات. يتم وضع جدول الحقائق في منتصف المخطط ويرتبط بجداول الأبعاد من خلال المراجع المناسبة. تظهر المقاييس في جدول الحقائق بخط عريض.

في بعض الأحيان ، ترتبط جداول الأبعاد بدورها بجداول أبعاد أخرى ، كما هو موضح في الشكل 3.3 ، من خلال عملية توحيد جزئي للبيانات ، من أجل تقليل استخدام الذاكرة. في المثال المعطى ، يرتبط جدول الأبعاد الذي يشير إلى الموقع بدوره بشكل هرمي بجدول الأبعاد الذي يحتوي على معلومات جغرافية. هذا يؤدي إلى مخطط ندفة الثلج snowflake scheme . يتضمن مستودع البيانات العديد من جداول الحقائق ، المترابطة بجداول الأبعاد ، المرتبطة بدورها بجداول الأبعاد الأخرى. يُطلق على النوع الأخير من المخطط ، الموضح في الشكل 3.4 ، مخطط المجرة Galaxy scheme.

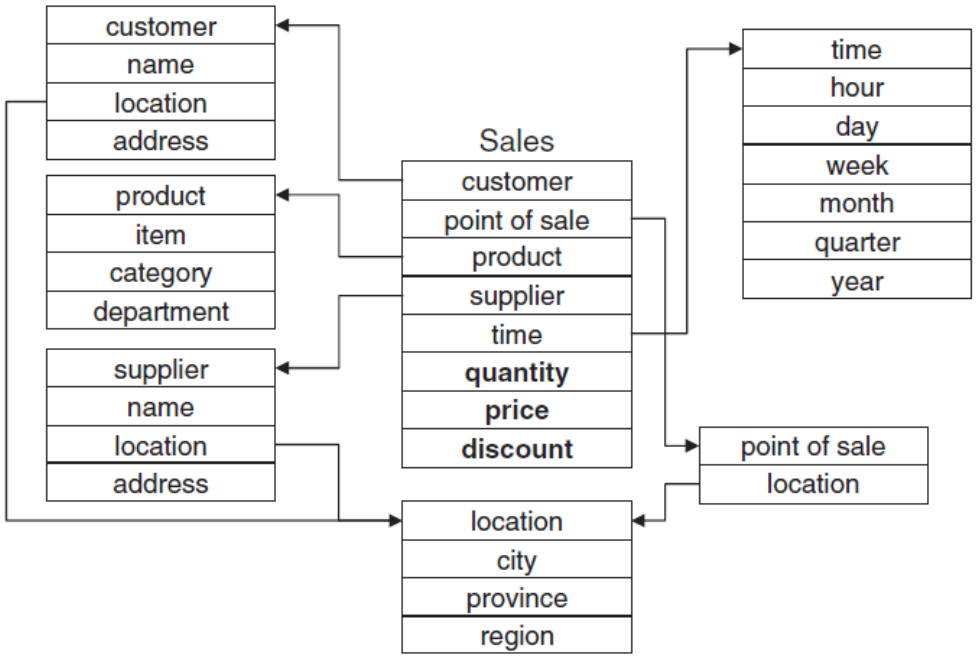
قد يتم تمثيل جدول الحقائق المرتبط بجداول أبعاد n بواسطة مكعب بيانات n-dimensional data cube حيث يتوافق كل محور مع بُعد. تعد المكعبات متعددة الأبعاد امتدادًا طبيعيًا لجداول البيانات ثنائية الأبعاد الشائعة ، والتي يمكن تفسيرها على أنها مكعبات ثنائية الأبعاد. على سبيل المثال ، ضع في اعتبارك جدول حقائق المبيعات الذي تم تطويره وفقًا للأبعاد الثلاثة لـ {الوقت ، المنتج ، المنطقة}. لنفترض أننا حددنا بعدين فقط للتحليل ، مثل {الوقت ، المنتج} ، بعد تعيين سمة المنطقة مسبقًا على طول القيم الثلاث {الولايات المتحدة الأمريكية ، وآسيا ، وأوروبا}. بهذه الطريقة نحصل على الجداول ثلاثية الأبعاد التي تتوافق فيها الصفوف مع أرباع السنة والأعمدة للمنتجات (انظر الجداول 3.3-3.5). المكعب الموضح في

الشكل 3.5 عبارة عن رسم توضيحي ثلاثي الأبعاد لنفس جدول حقائق المبيعات. يتم تمثيل البيانات الذرية من خلال 36 خلية يمكن الحصول عليها من خلال عبور جميع القيم الممكنة على طول الأبعاد الثلاثة: الوقت {Q1، Q2، Q3، Q4}، المنطقة {الولايات المتحدة الأمريكية، آسيا، أوروبا} والمنتج {TV، PC، DVD}.

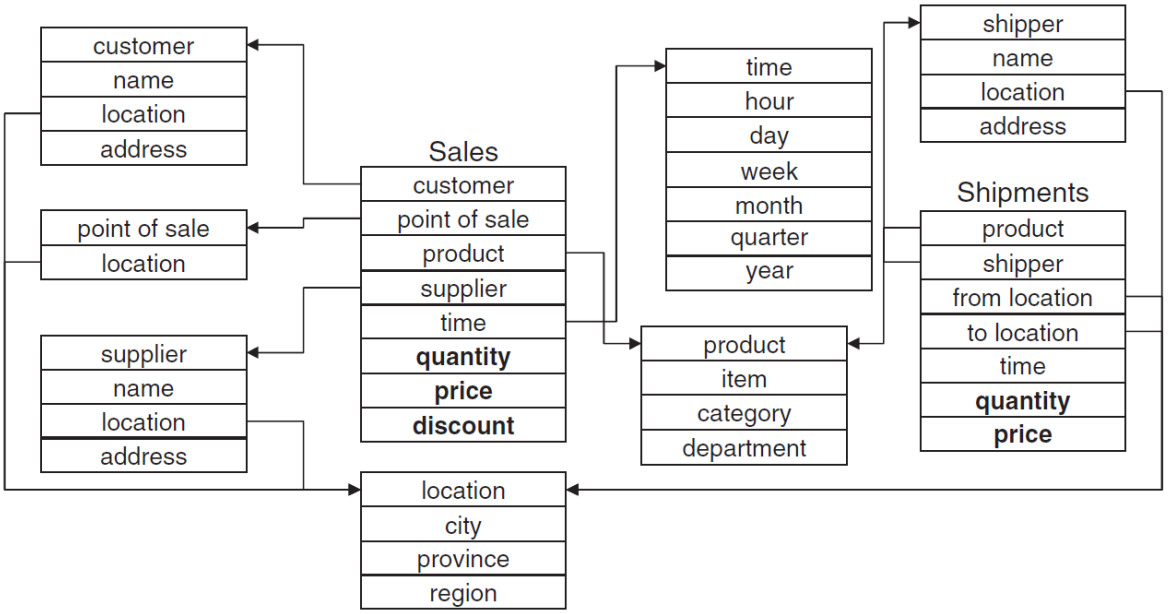
يمكن استكمال هذه الخلايا الذرية بـ 44 خلية تتوافق مع القيم الموجزة التي تم الحصول عليها من خلال الدمج على طول واحد أو أكثر من الأبعاد ، كما هو موضح بالمكعب في الشكل.



الشكل 3.2 مثال على مخطط نجمي



الشكل 3.3 مثال على مخطط ندفة الثلج



الشكل 3.4 مثال على مخطط المجرة

الجدول 3.3 عرض ثنائي الأبعاد لبيانات المبيعات في الولايات المتحدة الأمريكية

region = USA

	product		
time	TV	PC	DVD
Q1	980	546	165
Q2	765	456	231
Q3	879	481	192
Q4	986	643	203

الجدول 3.4 عرض ثنائي الأبعاد لبيانات المبيعات في آسيا

region = Asia

	product		
time	TV	PC	DVD
Q1	789	456	187
Q2	654	732	157
Q3	623	354	129
Q4	756	876	231

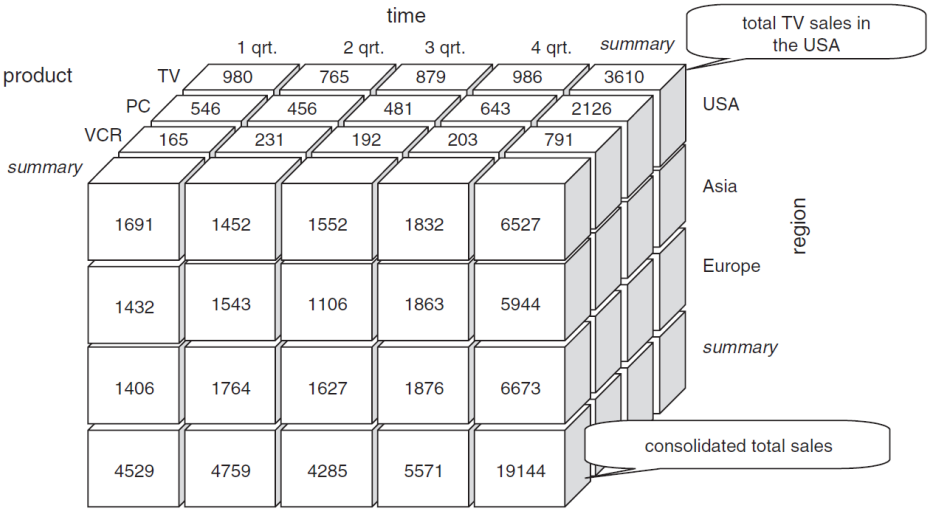
الجدول 3.5 عرض ثنائي الأبعاد لبيانات المبيعات في أوروبا

region = Europe

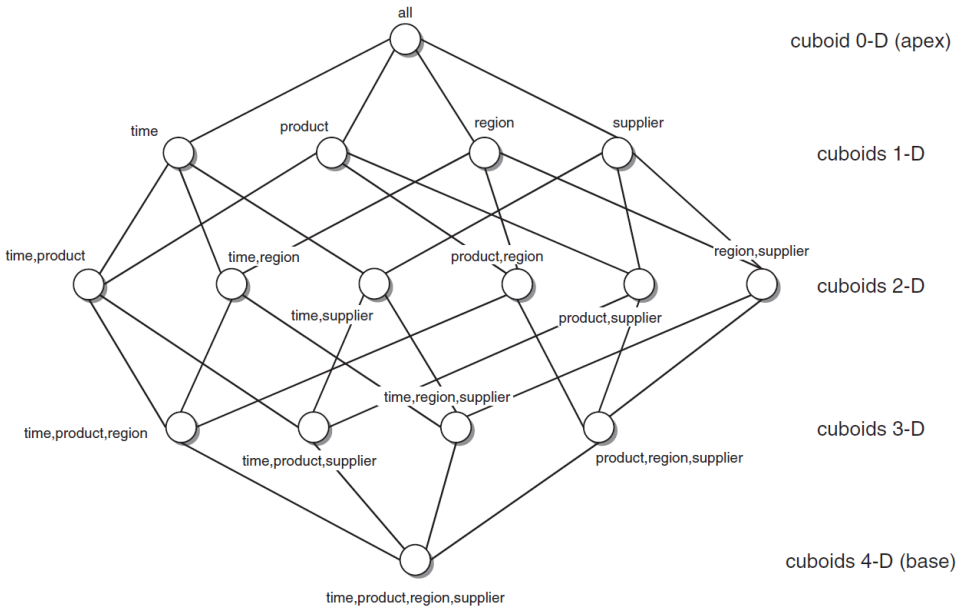
time	product		
	TV	PC	DVD
Q1	638	576	192
Q2	876	723	165
Q3	798	675	154
Q4	921	754	201

افترض أن جدول حقائق المبيعات يحتوي أيضًا على بُعد رابع يمثله الموردون. يشكل مكعب البيانات المقابل بنية في فضاء رباعي الأبعاد ، وبالتالي لا يمكن تمثيله بيانياً. ومع ذلك ، يمكننا الحصول على أربعة مناظر منطقية تتكون من مكعبات ثلاثية الأبعاد ، تسمى مكعبات ، داخل مكعب رباعي الأبعاد ، عن طريق تثبيت قيم ذات بعد واحد.

بشكل عام ، بدءًا من جدول حقائق مرتبط بجداول أبعاد n ، فهو من الممكن الحصول على شبكة من الأشكال المكعبة ، كل منها يتوافق مع مستوى مختلف من التوحيد على طول واحد أو أكثر من الأبعاد. هذا النوع من التجميع يكافئ في لغة الاستعلام المركبة (structured query language : SQL) مجموع استعلام مشتق من شرط تجميع group-by حسب. يوضح الشكل 3.6 الشبكة المكونة من الأشكال المكعبة التي تم الحصول عليها من مكعب البيانات المحدد على طول الأبعاد الأربعة {الوقت ، المنتج ، المنطقة ، المورد}.



الشكل 3.5 مثال على مكعب ثلاثي الأبعاد



الشكل 3.6 شبكية من شبه المكعبات مشتقة من مكعب رباعي الأبعاد

يسمى المكعب المرتبط بالبيانات الذرية ، والذي لا يعني بالتالي أي نوع من الدمج ، باسم متوازي المستطيلات القاعدية. في الطرف الآخر ، يتم تعريف القمة شبه المكعبة على أنها متوازي المستطيلات المقابلة للدمج على طول جميع الأبعاد ، وبالتالي فهي مرتبطة بالمجموع الكلي لمقياس الفائدة.

3.3.1 التسلسلات الهرمية للمفاهيم وعمليات OLAP

في كثير من الحالات ، تستند تحليلات OLAP إلى تسلسلات هرمية للمفاهيم لدمج البيانات وإنشاء طرق عرض منطقية على طول أبعاد مستودع البيانات. يحدد التسلسل الهرمي للمفهوم مجموعة من الخرائط من مستوى أدنى من المفاهيم إلى مستوى أعلى.

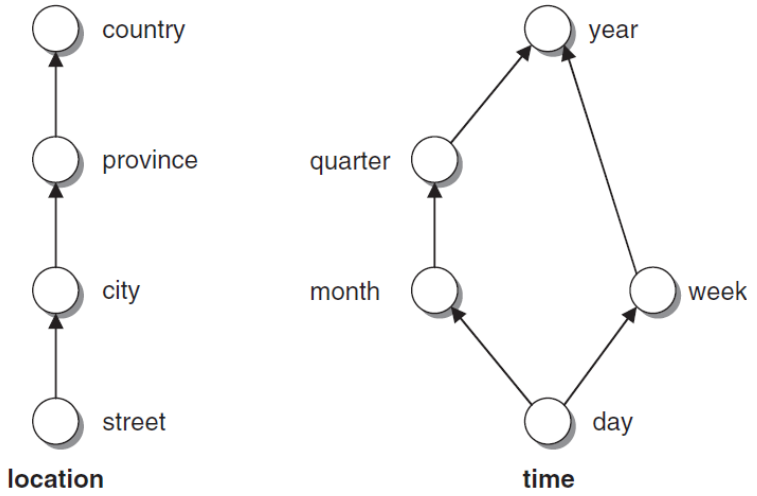
على سبيل المثال ، قد يُنشئ البعد {location} تسلسلاً هرمياً مرتبًا بالكامل ، كما هو موضح في الشكل 3.7 ، حيث يتم تطويره على طول علاقة {العنوان ، البلدية ، المقاطعة ، الدولة}. من ناحية أخرى ، فإن البعد الزمني ينشأ عن تسلسل هرمي مرتب جزئيًا ، كما هو موضح في الشكل 3.7.

يمكن تحديد أنواع التسلسل الهرمي المحددة مسبقًا في النظام الأساسي للبرنامج المستخدم لإنشاء وإدارة مستودع بيانات ، كما في حالة الأبعاد الموضحة في الشكل 3.7. بالنسبة للتسلسلات الهرمية الأخرى ، من الضروري أن يقوم المحللون بتحديد العلاقات بين المفاهيم بشكل صريح.

تُستخدم التسلسلات الهرمية للمفاهيم أيضًا لإجراء العديد من عمليات العرض المرئي visualization التي تتعامل مع مكعبات البيانات في مستودع البيانات.

1. **الدمج Roll-up أو consolidation.** تتكون هذه العملية، والتي تسمى أيضًا بعملية الحفر drill-up، من تجميع البيانات في المكعب ، والتي يمكن الحصول عليها بدلاً من ذلك بالطريقتين التاليتين.

- المضي قدمًا إلى مستوى أعلى على طول بعد واحد محدد عبر تسلسل هرمي للمفاهيم. على سبيل المثال ، بالنسبة للبعد {location} ، من الممكن الانتقال لأعلى من مستوى {city} إلى مستوى {مقاطعة} وتوحيد مقاييس الفائدة من خلال مجموع مشروط للمجموعة على جميع السجلات حيث تنتمي المدينة إلى نفس المقاطعة.



الشكل 3.7 التسلسلات الهرمية للمفاهيم

- التخفيض ببعد واحد. على سبيل المثال ، تؤدي إزالة البعد {time} إلى مقاييس موحدة من خلال المجموع على مدار جميع الفترات الزمنية الموجودة في مكعب البيانات.

2. **Roll-down.** هذه العملية والتي يشار إليها أيضًا باسم الحفر لأسفل ، هي العملية المعاكسة Roll-up. يسمح بالتنقل عبر مكعب البيانات من المعلومات المجمعة والموحدة إلى معلومات أكثر تفصيلاً. التأثير هو عكس النتيجة التي تم تحقيقها من خلال عملية الدمج. لذلك يمكن إجراء عملية الحفر بطريقتين.

- الانتقال إلى مستوى أدنى على طول التدرج الهرمي للبعد الفردي. على سبيل المثال ، في حالة البعد {location} ، من الممكن التحول من مستوى {المقاطعة} إلى المستوى {city} وتفصيل مقاييس الاهتمام على جميع السجلات حيث تنتمي المدينة إلى نفس المقاطعة.
- إضافة بعد واحد. على سبيل المثال ، يؤدي إدخال البعد {time} إلى فصل مقاييس الاهتمام على مدار جميع الفترات الزمنية الموجودة في مكعب البيانات.

3. **التقطيع Slice and Dice**. من خلال عملية التقطيع، يتم تحديد قيمة السمة وتثبيتها على طول بعد واحد. على سبيل المثال ، تم الحصول على الجدول 3.3 عن طريق تثبيت المنطقة بقيمة {الولايات المتحدة}. تحصل عملية التقطيع على مكعب في فضاء فرعي عن طريق تحديد عدة أبعاد في وقت واحد.
4. **المحور Pivot.** تنتج العملية المحورية ، التي يشار إليها أيضًا باسم التدوير ، دورانًا للمحاور ، وتبديل بعض الأبعاد للحصول على عرض مختلف لمكعب البيانات.

3.3.2 تجسيد مكعبات البيانات Materialization of cubes of data

قد تحتاج تحليلات OLAP التي طورها عمال المعرفة للوصول إلى المعلومات المرتبطة بالعديد من الأشكال المكعبة ، بناءً على الاستفسارات والتحليلات المحددة التي يتم إجراؤها. من أجل ضمان وقت استجابة مناسب ، قد يكون من المفيد تصميم مستودع بيانات حيث يتم حساب جميع (أو على الأقل جزء كبير من) قيم مقاييس الفائدة المرتبطة بجميع الأشكال المكعبة الممكنة. يسمى هذا النهج التجسيد الكامل full materialization للمعلومات المتعلقة بمكعبات البيانات.

لاحظ أنه عندما تكون التسلسلات الهرمية للمفاهيم مفقودة ، فمن الممكن تكوين n^2 من المكعبات متميزة من جميع التركيبات الممكنة للأبعاد n. إن وجود التسلسلات الهرمية على طول أبعاد مختلفة يجعل عدد الأشكال المكعبة المتميزة أكبر.

إذا كان L_i يشير إلى عدد المستويات الهرمية المرتبطة بالبعد i، فبالنسبة لمكعب بيانات n-dimensional ، من الممكن حساب العدد الكامل للمكعبات ، بأعطاء التالي:

$$T = \prod_{i=1}^{n} (L_i + 1)$$

على سبيل المثال ، إذا كان مكعب البيانات يشتمل على 5 أبعاد ، وإذا كان كل من هذه الأبعاد يتضمن 3 مستويات متدرجة ، فإن عدد الأشكال المكعبة يساوي 45 = 210 ≈ 103. من الواضح أن التجسيد الكامل للمكعبات المستطيلة لجميع المكعبات المرتبطة بجداول الحقائق الخاصة بمستودع البيانات من شأنه أن يفرض متطلبات تخزين يصعب الحفاظ عليها بمرور الوقت ، مع الأخذ في الاعتبار معدل جمع السجلات الجديدة.

لجميع الأسباب المذكورة أعلاه ، من الضروري تحقيق توازن بين الحاجة إلى الوصول السريع إلى المعلومات ، مما قد يشير إلى التجسيد الكامل للمكعبات ، والحاجة إلى الحفاظ على استخدام الذاكرة ضمن حدود معقولة.

نتيجة لذلك ، يجب تنفيذ التجسيد الوقائي فقط لتلك المكعبات التي يتم الوصول إليها بشكل متكرر ، بينما بالنسبة للآخرين ، يجب إجراء الحساب عند الطلب فقط عند إجراء استفسارات فعلية تطلب المعلومات المرتبطة. يشار إلى هذا النهج الأخير على أنه تجسيد جزئي للمعلومات المتعلقة بمكعبات البيانات.

3.4 الملاحظات والقراءات

تم توفير المزيد من الدراسات المتعمقة لمخازن البيانات ، بما في ذلك الجوانب التقنية التي تم حذفها عمدًا من هذا الفصل ، بواسطة ، Kimball et al. (1998) ، Kimball (1996) Kimball and Ross (2002 ، Inmon (2002) ، (Dyche (2000 و Nemati et al. (2002)).

الجزء الثاني

النماذج والطرق الرياضية

الفصل الرابع

النماذج الرياضية لصنع القرار

في الفصول السابقة أكدنا على الدور الحاسم الذي تلعبه النماذج الرياضية في تطوير بيئات ذكاء الأعمال وأنظمة دعم القرار التي تهدف إلى توفير الدعم الفعال للعاملين في مجال المعرفة. سنركز في هذا الفصل على الخصائص الرئيسية المشتركة بين النماذج الرياضية المختلفة المضمنة في أنظمة ذكاء الأعمال.

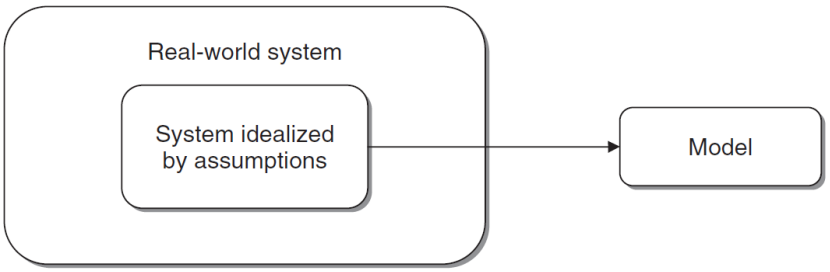
سنقوم أيضًا بتطوير تصنيف لفئات النماذج الأكثر شيوعًا ، وتحديد مجال التطبيق السائد لكل منها.

4.1 هيكل النماذج الرياضية

تم تطوير النماذج الرياضية واستخدامها في العديد من مجالات التطبيق ، بدءًا من الفيزياء إلى الهندسة المعمارية ومن الهندسة إلى الاقتصاد.

تختلف النماذج المعتمدة في السياقات المختلفة اختلافًا جوهريًا من حيث هيكلها الرياضي. ومع ذلك ، من الممكن تحديد بعض الميزات الأساسية المشتركة بين معظم النماذج.

بشكل عام ، النموذج model هو تجريد انتقائي لنظام حقيقي. بمعنى آخر ، تم تصميم النموذج لتحليل وفهم السلوك التشغيلي لنظام حقيقي من وجهة نظر مجردة ، والذي يتضمن فقط العناصر التي تعتبر ذات صلة لغرض التحقيق الذي تم إجراؤه. في هذا الصدد ، يجدر اقتباس كلمات أينشتاين عن تطوير النموذج: "كل شيء يجب أن يكون بسيطًا قدر الإمكان ، ولكن ليس أبسط". يوضح الشكل 4.1 بمصطلحات رسومية تعريف النموذج.



الشكل 4.1 النموذج هو تجريد انتقائي للواقع

لقد تحول التطور العلمي والتكنولوجي إلى نماذج رياضية من أنواع مختلفة للتمثيل المجرد للأنظمة الحقيقية. على سبيل المثال ، انظر إلى التجربة الفكرية (Gendanken experiment) التي انتشرت في الفيزياء في بداية القرن العشرين ، والتي تضمنت بناء نموذج عقلي لظاهرة معينة والتحقق من صحتها من خلال تخيل النتائج التي تسببها التعديلات الافتراضية في النموذج نفسه. التشابه واضح بين هذا النموذج المفاهيمي وتحليلات ماذا لو what if analysis التي يمكن إجراؤها بسهولة باستخدام جدول بيانات بسيط للعثور على إجابة لأسئلة مثل: إعطاء نموذج لحساب ميزانية الشركة ، كيف تتأثر التدفقات النقدية بـ تغيير في شروط الدفع ، مثل 90 يومًا مقابل 60 يومًا ، للفواتير الصادرة لصالح العملاء الرئيسيين؟ وفقًا لخصائصها ، يمكن تقسيم النماذج إلى أيقونات وتناظرية ورمزية.

- **الأيقونية Iconic**. النموذج الأيقوني هو تمثيل مادي لنظام حقيقي ، يتم تقليد سلوكه لغرض التحليل. نموذج مصغر لحي مدينة جديد هو مثال على نموذج مبدع.
- **التناظرية Analogical**. النموذج القياسي هو أيضًا تمثيل مادي ، على الرغم من أنه يقلد السلوك الحقيقي عن طريق القياس بدلاً من التكرار. يُعد نفق الرياح الذي تم إنشاؤه للتحقق من الخصائص الديناميكية الهوائية للسيارة مثالاً على نموذج تماثلي يهدف إلى تمثيل التقدم الفعلي للمركبة على الطريق.
- **الرمزية Symbolic**. النموذج الرمزي ، مثل النموذج الرياضي ، هو تمثيل تجريدي لنظام حقيقي. الغرض منه هو وصف سلوك النظام من خلال سلسلة من المتغيرات الرمزية والمعلمات العددية والعلاقات الرياضية.

أن أنظمة ذكاء الأعمال ، وبالتالي النماذج المعروضة في هذا الكتاب ، هي نماذج رمزية.

يتعلق تمييز آخر ذي صلة بالطبيعة الاحتمالية للنماذج ، والتي يمكن أن تكون إما عشوائية stochastic أو حتمية deterministic.

- **العشوائية stochastic** . في النموذج العشوائي ، تمثل بعض معلومات الإدخال أحداثًا عشوائية وبالتالي تتميز بتوزيع احتمالي ، والذي بدوره يمكن تعيينه أو عدم معرفته. النماذج التنبؤية ، التي سيتم وصفها بدقة في الفصول التالية ، وكذلك نماذج خط الانتظار ، المذكورة بإيجاز أدناه في هذا الفصل ، هي أمثلة على النماذج العشوائية.

- **محددة - حتمية deterministic.** يُطلق على النموذج اسم محدد او حتمي عندما يُفترض أن تكون جميع بيانات الإدخال معروفة مسبقًا وعلى وجه اليقين. نظرًا لأن هذا الافتراض نادرًا ما يتم تحقيقه في الأنظمة الحقيقية ، يلجأ المرء إلى النماذج الحتمية عندما تكون المشكلة المطروحة معقدة بدرجة كافية وتكون أي عناصر عشوائية ذات أهمية محدودة.

لاحظ ، مع ذلك ، أنه حتى بالنسبة للنماذج الحتمية ، قد يتم تخفيف فرضية معرفة البيانات على وجه اليقين. تسمح تحليلات الحساسية والسيناريوهات ، بالإضافة إلى تحليل ماذا لو what if analysis، بتقييم قوة القرارات المثلى للتغيرات في معلمات الإدخال.

هناك تمييز آخر يتعلق بالبعد الزمني في النموذج الرياضي ، والذي يمكن أن يكون إما ثابتًا أو ديناميكيًا.

- **ثابتة Static.** تعتبر النماذج الثابتة نظامًا معينًا وعملية صنع القرار ذات الصلة ضمن مرحلة زمنية واحدة. على سبيل المثال ، يحدد نموذج التحسين الموضح في القسم 14.2.7 خطة مثالية لتوزيع البضائع في إطار زمني محدد.
- **متحرك Dynamic.** تنظر النماذج الديناميكية إلى نظام معين من خلال عدة مراحل زمنية تتوافق مع سلسلة من القرارات. في كثير من الحالات ، يتم تقسيم البعد الزمني إلى فواصل زمنية منفصلة لفترة محددة مسبقًا: الدقائق والساعات والأيام والأسابيع والأشهر والسنوات هي أمثلة على التقسيمات الفرعية المنفصلة لمحور الوقت. النماذج الديناميكية للوقت المنفصل Discrete-time، والتي تسود بشكل كبير في تطبيقات ذكاء الأعمال ، تراقب حالة النظام فقط في بداية أو في نهاية الفواصل الزمنية المنفصلة. تأخذ النماذج الديناميكية المستمرة للوقت Continuous-time في الاعتبار تسلسلًا مستمرًا للفترات على محور الوقت.

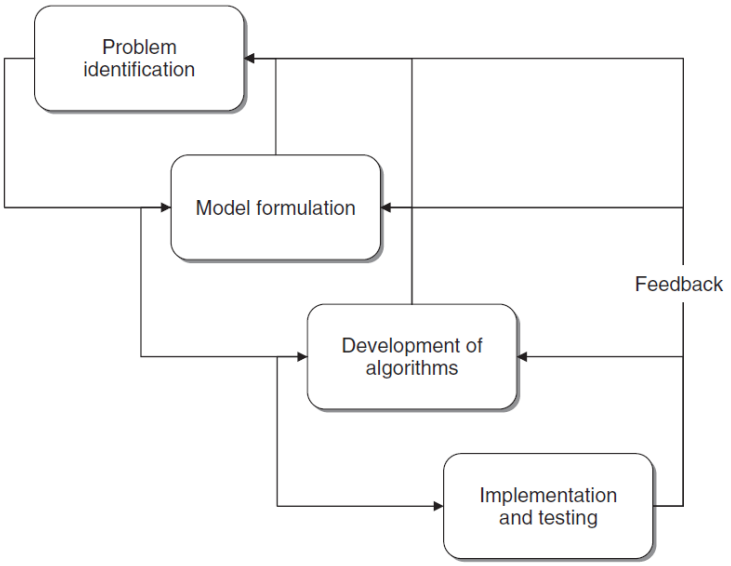
4.2 تطوير نموذج

من الممكن تقسيم تطوير نموذج رياضي لصنع القرار إلى أربع مراحل أولية ، كما هو موضح في الشكل 4.2. يتضمن الشكل أيضًا آلية تغذية مرتدة feedback تأخذ في الاعتبار إمكانية إجراء تغييرات ومراجعات للنموذج.

تحديد المشكلة Problem identification

أولاً وقبل كل شيء ، يجب تحديد المشكلة المطروحة بشكل صحيح. يجب تحليل الأعراض الحرجة الملحوظة وتفسيرها من أجل صياغة فرضيات للتحقيق. على سبيل المثال ، قد يمثل مستوى المخزون المرتفع للغاية ، والذي يقابل معدل دوران المخزون المفرط ، أحد أعراض الشركة التي تصنع السلع الاستهلاكية. لذلك من الضروري فهم سبب المشكلة ، بناءً على رأي مديري الإنتاج.

في هذه الحالة ، قد تكون خطة الإنتاج غير الفعالة هي سبب تراكم المخزون.



الشكل 4.2 مراحل تطوير النماذج الرياضية لصنع القرار

صياغة النموذج Model formulation

بمجرد تحديد المشكلة المراد تحليلها بشكل صحيح ، يجب توجيه الجهد نحو تحديد نموذج رياضي مناسب لتمثيل النظام. هناك عدد من العوامل التي تؤثر على اختيار النموذج وتؤثر عليه ، مثل الأفق الزمني time horizon ومتغيرات القرار decision variables ومعايير التقييم evaluation criteria والمعلمات العددية numerical parameters والعلاقات الرياضية mathematical relationships.

- **أفق زمني time horizon.** عادة ما يتضمن النموذج بعدا زمنيا. على سبيل المثال ، لصياغة خطة إنتاج تكتيكية على المدى المتوسط ، من الضروري تحديد معدل الإنتاج

لكل أسبوع في السنة ، بينما لاشتقاق جدول تشغيلي ، يلزم تعيين المهام لكل خط إنتاج لكل يوم من أيام السنة. أسبوع. كما يمكننا أن نرى ، فإن الفترة الزمنية التي يتم أخذها في الاعتبار في النموذج ، وكذلك طول الفواصل الزمنية الأساسية ، قد تختلف اعتمادًا على المشكلة المحددة التي تم أخذها في الاعتبار.

- **معيار التقييم evaluation criteria**. يجب تحديد مؤشرات الأداء المناسبة والقابلة للقياس من أجل وضع معيار لتقييم ومقارنة القرارات البديلة. قد تتخذ هذه المؤشرات أشكالًا مختلفة في كل تطبيق مختلف ، وقد تشمل العوامل التالية:
  - التكاليف والمكافآت النقدية.
  - فعالية ومستوى الخدمة.
  - جودة المنتجات والخدمات.
  - مرونة ظروف التشغيل.
  - الموثوقية في تحقيق الأهداف.
- **متغيرات القرار decision variables**. يجب بعد ذلك تحديد المتغيرات الرمزية التي تمثل قرارات بديلة. على سبيل المثال ، إذا كانت المشكلة تتكون من صياغة خطة إنتاج تكتيكية على المدى المتوسط ، فيجب أن تعبر متغيرات القرار عن أحجام الإنتاج لكل منتج ولكل عملية ولكل فترة من أفق التخطيط.
- **المعلمات العددية numerical parameters**. من الضروري أيضًا تحديد وتقدير جميع المعلمات الرقمية التي يتطلبها النموذج بدقة. في مثال تخطيط الإنتاج ، يجب معرفة السعة المتاحة مسبقًا لكل عملية ، بالإضافة إلى معاملات امتصاص السعة لكل مجموعة من المنتجات والعمليات.
- **العلاقات الرياضية mathematical relationships**. تمثل الخطوة الأخيرة في صياغة النموذج في تحديد العلاقات الرياضية بين متغيرات القرار والمعلمات العددية ومؤشرات الأداء المحددة خلال المراحل السابقة. في بعض الأحيان قد تكون هذه العلاقات حتمية بشكل حصري ، بينما في حالات أخرى يكون من الضروري تقديم علاقات احتمالية.

في هذه المرحلة ، يجب النظر بعناية في المفاضلة بين دقة التمثيل المحقق من خلال النموذج وتعقيد الحل.
قد يكون من المفيد أكثر على المستوى العملي اعتماد نموذج يضحي ببعض الجوانب الهامشية للواقع في تمثيل النظام ولكنه يسمح بحل فعال ومرونة أكبر في ضوء التطورات المستقبلية المحتملة.

تطوير الخوارزميات Development of algorithms

بمجرد تحديد نموذج رياضي ، سيرغب المرء بطبيعة الحال في المضي قدمًا في حله لتقييم القرارات واختيار البديل الأفضل. بمعنى آخر ، يجب تحديد خوارزمية الحل ويجب تطوير أداة برمجية تتضمن طريقة الحل أو الحصول عليها. يجب أن يمتلك المحلل المسؤول عن صياغة النموذج معرفة شاملة بأساليب الحل الحالية وخصائصها.

التنفيذ والاختبار Implementation and test

عندما يتم تطوير النموذج بالكامل ، يتم أخيرًا تنفيذه واختباره واستخدامه في مجال التطبيق. من الضروري أيضًا أن يتم تقييم صحة البيانات والمعلمات الرقمية التي تم إدخالها في النموذج بشكل أولي. تأتي هذه البيانات عادةً من مستودع بيانات أو سوق بيانات تم إعداده مسبقًا. بمجرد الحصول على النتائج الرقمية الأولى باستخدام إجراء الحل المبتكر ، يجب التحقق من صحة النموذج من خلال تقديم استنتاجاته إلى رأي صانعي القرار وغيرهم من الخبراء في مجال التطبيق. ينبغي أن يؤخذ عدد من العوامل في الاعتبار في هذه المرحلة:

- معقولية واحتمالية الاستنتاجات التي تم تحقيقها
- اتساق النتائج عند القيم القصوى للمعلمات العددية
- استقرار النتائج عند إدخال تغييرات طفيفة في معلمات الإدخال.

4.3 فئات النماذج

هناك عدة فئات من النماذج الرياضية لصنع القرار ، والتي بدورها يمكن حلها من خلال عدد من تقنيات الحلول البديلة. كل فئة نموذجية هي الأنسب لتمثيل أنواع معينة من عمليات صنع القرار.

في هذا القسم سوف نغطي الفئات الرئيسية للنماذج الرياضية لصنع القرار ، بما في ذلك:

1. النماذج التنبؤية.
2. التعرف على الأنماط ونماذج التعلم.
3. نماذج التحسين.
4. نماذج إدارة المشروع.
5. نماذج تحليل المخاطر.
6. نماذج خط الانتظار.

1. النماذج التنبؤية Predictive models

تتطلب نسبة كبيرة من النماذج المستخدمة في أنظمة ذكاء الأعمال ، مثل نماذج التحسين ، بيانات إدخال تتعلق بالأحداث المستقبلية.

على سبيل المثال ، تحدد نتائج الأحداث العشوائية الطلب المستقبلي على منتج أو خدمة ، وتطوير سيناريوهات جديدة للابتكار التكنولوجي ومستوى الأسعار والتكاليف. نتيجة لذلك ، تلعب النماذج التنبؤية دورًا أساسيًا في أنظمة ذكاء الأعمال ، حيث يتم وضعها بشكل منطقي في المقدمة فيما يتعلق بالنماذج الرياضية الأخرى ، وبشكل أعم ، في عملية صنع القرار بأكملها.

تسمح التنبؤات بإدخال معلومات الإدخال في عمليات صنع القرار المختلفة ، الناشئة في الإستراتيجية والبحث والتطوير والإدارة والرقابة والتسويق والإنتاج والخدمات اللوجستية. في الأساس ، تستخدم جميع وظائف الإدارات في المؤسسة بعض المعلومات التنبؤية لتطوير عملية صنع القرار ، على الرغم من أنها تسعى إلى تحقيق أهداف مختلفة.

يمكن تقسيم النماذج التنبؤية إلى فئتين رئيسيتين. الغرض من النماذج التوضيحية explanatory models هو التحديد الوظيفي للعلاقة المحتملة بين متغير تابع ومجموعة من السمات المستقلة. تنتمي نماذج الانحدار Regression models، الموصوفة في الفصل 8 ، إلى

هذه الفئة بالإضافة إلى نماذج التصنيف classification models الموضحة في الفصل 10 ، على الرغم من أنه سيكون من الأنسب وضع هذه الأخيرة جنبًا إلى جنب مع التعرف على الأنماط ونماذج التعلم ، التي تمت تغطيتها في الفقرة التالية. الغرض من نماذج السلاسل الزمنية time series models، الموصوفة في الفصل 9 ، هو التحديد الوظيفي لأي نمط زمني معبر عنه بسلسلة زمنية من الملاحظات المشار إليها في نفس المتغير العددي.

2. التعرف على الأنماط ونماذج تعليم الماكنة Pattern recognition and machine learning models

بالمعنى الواسع ، الغرض من التعرف على الأنماط ونظرية التعلم هو فهم الآليات التي تنظم تطور الذكاء ، والتي تُفهم على أنها القدرة على استخراج المعرفة من التجربة السابقة من أجل تطبيقها في المستقبل. يمكن استخدام النماذج الرياضية للتعلم لتطوير خوارزميات فعالة يمكنها أداء مثل هذه المهمة. وقد أدى ذلك إلى آلات ذكية قادرة على التعلم من الملاحظات السابقة واشتقاق قواعد جديدة للمستقبل ، تمامًا كما يستطيع العقل البشري أن يفعل بفاعلية كبيرة بسبب الآليات المعقدة التي تم تطويرها وضبطها في مسار التطور.

إلى جانب الاهتمام النظري الجوهري ، يتم تطبيق الأساليب الرياضية للتعلم في العديد من المجالات ، مثل التعرف على الصور والأصوات والنصوص ؛ التشخيص البيولوجي والطبي. التسويق العلائقي ، لتقسيم العملاء وتنميطهم ؛ مراقبة عملية التصنيع ؛ تحديد الحالات الشاذة والكشف عن الاحتيال.

النماذج الرياضية للتعلم لها هدفان أساسيان. الغرض من نماذج التفسير هو تحديد الأنماط المنتظمة في البيانات والتعبير عنها من خلال قواعد ومعايير يسهل فهمها. تساعد نماذج التنبؤ على التنبؤ بالقيمة التي سيفترضها متغير عشوائي معين في المستقبل ، بناءً على قيم بعض المتغيرات المرتبطة بكيانات قاعدة البيانات ، كما هو الحال بالنسبة للنماذج التوضيحية. بناءً على وجود أو عدم وجود سمة مستهدفة ، قد يتم الإشراف على عملية التعلم أو عدم الإشراف عليها supervised or unsupervised. في الحالة الأولى ، تعبر السمة الهدف لكل سجل إما عن فئة العضوية أو كمية قابلة للقياس. تنتمي نماذج التصنيف والانحدار إلى هذه الفئة. في الحالة الثانية ، لا توجد سمة مستهدفة ، وبالتالي فإن الغرض من التحليل هو تحديد أوجه الانتظام والتشابه

والاختلاف في البيانات. من الممكن أيضًا اشتقاق قواعد الارتباط ، كما هو موضح في الفصل 11. وبدلاً من ذلك ، يمكن تحديد مجموعات من السجلات ، تسمى المجموعات ، والتي تتميز بالتشابه داخل كل مجموعة وبسبب الاختلاف بين عناصر المجموعات المميزة ، كما هو موضح في الفصل 12.

3. نماذج التحسين Optimization models

يمكن صياغة العديد من عمليات صنع القرار التي تواجهها الشركات أو المنظمات المعقدة وفقًا للإطار التالي: نظرًا للمشكلة المطروحة ، يحدد صانع القرار مجموعة من القرارات الممكنة feasible decisions ويضع معيارًا لتقييم ومقارنة الخيارات البديلة ، مثل التكاليف المالية أو المكافآت. في هذه المرحلة ، يجب على صانع القرار تحديد القرار الأمثل optimal وفقًا لمعيار التقييم المحدد ، أي الاختيار المقابل للحد الأدنى للتكلفة أو الحد الأقصى للمكافأة. يحدد النموذج المفاهيمي المحدد فئة واسعة وشائعة من النماذج الرياضية لصنع القرار ، ممثلة بنماذج التحسين.

بشكل عام ، تظهر نماذج التحسين بشكل طبيعي في عمليات صنع القرار حيث يجب تخصيص مجموعة من الموارد المحدودة بأكثر الطرق فعالية للكيانات المختلفة. قد تكون هذه الموارد موظفين أو عمليات إنتاج أو مواد خام أو مكونات أو عوامل مالية. من بين مجالات التطبيق الرئيسية التي تتطلب التخصيص الأمثل للموارد نجد:

- التخطيط اللوجستي والإنتاج.
- التخطيط المالي؛
- تخطيط ورديات العمل.
- تخطيط حملة التسويق.
- تحديد السعر.

تمثل نماذج التحسين الرياضي Mathematical optimization فئة كبيرة إلى حد ما من مشاكل التحسين التي يتم اشتقاقها عندما يكون الهدف من عملية اتخاذ القرار هو دالة لمتغيرات القرار ، ويمكن التعبير عن المعايير التي تصف القرارات الممكنة من خلال مجموعة من المساواة الرياضية وعدم المساواة في متغيرات القرار . توفر نماذج التحسين الرياضي فرصًا كبيرة للتطبيق ، كما سنرى في الفصول التالية ، نظرًا للمرونة العالية لتركيباتها وتوافر طرق حل فعالة.

في ضوء بنية الوظيفة الموضوعية والقيود ، قد تتخذ نماذج التحسين أشكالًا مختلفة:

- التحسين الخطي linear optimization.
- تحسين عدد صحيح integer optimization.
- التحسين المحدب convex optimization.
- تحسين الشبكة network optimization.
- التحسين متعدد الأهداف multiple-objective optimization.

4. نماذج إدارة المشروع Project management models

المشروع عبارة عن مجموعة معقدة من الأنشطة المترابطة التي يتم تنفيذها سعياً وراء هدف محدد ، والذي قد يمثل مصنعًا صناعيًا أو مبنى أو نظام معلومات أو منتجًا جديدًا أو هيكلًا تنظيميًا جديدًا ، اعتمادًا على مجالات التطبيق المختلفة . يتطلب تنفيذ المشروع عملية تخطيط ومراقبة للأنشطة المترابطة وكذلك الموارد البشرية والتقنية والمالية اللازمة لتحقيق الهدف النهائي. تعتمد أساليب إدارة المشروع على مساهمات مختلف التخصصات ، مثل تنظيم الأعمال وعلم النفس السلوكي وبحوث العمليات.

تلعب النماذج الرياضية لصنع القرار دورًا مهمًا في طرق إدارة المشروع. على وجه الخصوص ، تُستخدم نماذج الشبكة لتمثيل الأنشطة المكونة للمشروع وعلاقات الأسبقية فيما بينها.

تسمح هذه النماذج بتحديد وقت تنفيذ المشروع الإجمالي ، بافتراض معرفة حتمية بمدة كل نشاط.

من ناحية أخرى ، تُستخدم النماذج العشوائية ، التي يشار إليها عادةً باسم تقنيات تقييم ومراجعة المشروع (project evaluation and review techniques, PERT) ، لاشتقاق أوقات التنفيذ عند وضع افتراضات عشوائية فيما يتعلق بمدة الأنشطة ، ممثلة بمتغيرات عشوائية.

أخيرًا ، تسمح فئات مختلفة من نماذج التحسين بتوسيع التحليل ليشمل المشكلة المعقدة المتمثلة في التخصيص الأمثل لمجموعة من الموارد المحدودة بين أنشطة المشروع في ضوء تكاليف التنفيذ والأوقات.

5. نماذج تحليل المخاطر Risk analysis models

يمكن وصف بعض مشكلات القرار وفقًا للنموذج المفاهيمي التالي: يتعين على صانع القرار الاختيار من بين عدد من البدائل المتاحة ، مع وجود معلومات غير مؤكدة بشأن التأثيرات التي قد تحدث هذه الخيارات في المستقبل.

على سبيل المثال ، افترض أن الإدارة العليا ترغب في تقييم البدائل المختلفة من أجل زيادة الطاقة الإنتاجية للشركة. من ناحية أخرى ، قد تقوم الشركة ببناء مصنع جديد يوفر كفاءة تشغيل عالية ويتطلب تكلفة استثمارية عالية. من ناحية أخرى ، قد تقوم بتوسيع مصنع موجود باستثمار أقل ولكن بتكاليف تشغيل أعلى. أخيرًا ، قد تتعاقد من الباطن مع أطراف ثالثة خارجية على جزء من إنتاجها: في هذه الحالة ، يتم تقليل تكلفة الاستثمار إلى الحد الأدنى ولكن تكاليف التشغيل هي الأعلى بين البدائل المتاحة.

من الواضح ، في هذه الحالة ، أن تأثيرات الخيارات المختلفة تتأثر بشدة بالأحداث العشوائية المستقبلية. على وجه الخصوص ، فإن المستوى العالي من الطلب المستقبلي يجعل بناء مصنع جديد مفيدًا ، بينما تميل مستويات الطلب المنخفضة إلى تفضيل خيار التعاقد من الباطن. عند المستوى المتوسط للطلب ، قد يكون توسيع مصنع قائم مناسبًا.

ومع ذلك ، فإن صانع القرار مجبر على اتخاذ قرار قبل أن يعرف على وجه اليقين مستوى الطلب في المستقبل. في أحسن الأحوال ، قد تحصل على بعض المعلومات العشوائية فيما يتعلق باحتمالية حدوث أحداث مستقبلية من خلال إجراء بعض أبحاث السوق.

في حالات من هذا النوع ، قد يكون الدعم المنهجي الذي تقدمه نماذج تحليل المخاطر ، المستندة أساسًا إلى نظريات بايزي ونظريات المنفعة ، مفيدًا للغاية. في الواقع ، يتم استخدام هذه الفئة من النماذج بنجاح في العديد من مجالات التطبيق ، مثل الاستثمار التكنولوجي ، وتصميم المنتجات الجديدة ، والبحث والتطوير ، والاستثمار المالي والعقاري.

6. نماذج خط الانتظار Waiting line models

الغرض من نظرية خط الانتظار هو التحقيق في ظواهر الازدحام التي تحدث عندما يكون الطلب على الخدمة وتقديمها عشوائيًا بطبيعته. إذا لم تكن أوقات وصول العملاء ومدة الخدمة

معروفة مسبقًا بطريقة حتمية ، فقد تنشأ تعارضات بين العملاء في استخدام الموارد المشتركة المحدودة. نتيجة لذلك ، يضطر بعض العملاء إلى الانتظار في طابور.

من الناحية التخطيطية ، يتكون نظام خط الانتظار من ثلاثة مكونات رئيسية: مصدر يولد عملية عشوائية تصل من خلاله الكيانات ، التي يشار إليها أيضًا باسم العملاء ، إلى موقع معين للحصول على خدمة ؛ مجموعة من الموارد التي تقدم الخدمة ؛ منطقة انتظار قادرة على استقبال الكيانات التي لا يمكن تلبية طلباتها على الفور.

تسمح نماذج خط الانتظار بتقييم أداء النظام بمجرد تحديد هيكله ، وبالتالي فهي مفيدة في الغالب في مرحلة تصميم النظام. في الواقع ، من أجل تحديد القيم المناسبة للمعلمات التي تميز النظام الجديد ، يتم النظر في العوامل الاقتصادية ذات الصلة ، والتي تعتمد على مستوى الخدمة الذي يجب أن يضمنه النظام عند التشغيل في الظروف المثلى. بتعبير أدق ، يأخذ النموذج في الاعتبار تكلفة تلبية الطلبات ، والتي تزداد مع زيادة مستوى الخدمة ، والتكلفة الناتجة عن أوقات انتظار العملاء ، والتي تقل مع انخفاض مستوى الخدمة المقدمة. نظرًا لأن العملاء يأملون في فترات انتظار أقصر بينما يكون المزود مهتمًا بخفض تكلفة تقديم الخدمة ، يتم تحديد هيكل النظام بطريقة للحصول على المفاضلة المثلى بين تكاليف الخدمة وتكاليف خط الانتظار. بمعنى آخر ، يمكن العثور على مستوى الخدمة الأمثل ، والذي بدوره يحدد الهيكل المثالي للنظام ، عند الحد الأدنى من نقطة المنحنى التي تعبر عن مجموع نوعي التكلفة.

المكونات الرئيسية لنظام خط الانتظار هي السكان population ومعالجة الوصول arrivals process ومعالجة الخدمة service process وعدد المحطات number of stations وقواعد خط الانتظار waiting line rules. يمثل **السكان** ، الذي يمكن أن يكون محدودًا أو غير محدود ، المصدر الذي يتم من خلاله جذب العملاء المحتملين والذي يعودون إليه بمجرد تلقي الخدمة المطلوبة. تصف **معالجة الوصول** كيفية وصول العملاء إلى نقطة دخول النظام. بشكل عام ، هذه عملية عشوائية موصوفة من خلال التوزيع الاحتمالي لأوقات ما بين الوصول ، أي الفترات الزمنية بين وصول عميلين متتاليين. تصف **معالجة الخدمة** كيفية تلبية مقدمي الخدمة لطلبات العملاء المنتظرين في الطابور. هذه بدورها عملية عشوائية ، يتم تحديدها من خلال التوزيع الاحتمالي لوقت الخدمة ، أي مقدار الوقت الذي يقضيه العملاء مع الموارد التي تقدم الخدمة. يعد **عدد المحطات** الموجودة وعدد مقدمي الخدمة المعينين لكل محطة معلمات

إضافية ذات صلة لنظام خط الانتظار. تصف **قواعد الانتظار** الترتيب الذي يتم به إخراج العملاء من الخط ليتم قبولهم في الخدمة. أخيرًا ، يتم لعب الدور الأساسي من خلال مخططات الأولوية التي يتم فيها تعيين مستوى الأولوية لكل عميل. ثم يتم تقديم العميل ذي الأولوية القصوى قبل انتظار جميع العملاء الآخرين في الطابور.

4.4 ملاحظات وقراءات

للمواضيع التي ستتم مناقشتها في الفصول التالية ، مثل التعرف على الأنماط والنماذج التنبؤية ، راجع قسم القراءة المقترح في الفصول المعنية. لمزيد من المعلومات المتعمقة حول فئات معينة من النماذج ، انظر: (2003) Bertsekas ، (2002) Winston and Venkataramanan (2008 and Vercellis) حول التحسين الرياضي ، و 1994) .Nemhauser et al) بشأن التحسين بشكل عام ؛ المغربي (1978) حول نماذج الشبكة لإدارة المشاريع. Clemen 1997)) و 1999) Skinner) حول تحليل المخاطر ؛ 1975) Kleinrock) عن نظرية خط الانتظار. تتعامل 1995) Keys) مع الجوانب المعرفية المتضمنة في بحوث العمليات.

الفصل الخامس

التنقيب عن البيانات Data Mining

كما لوحظ في الفصول السابقة ، أتاحت التقنيات المتطورة لجمع المعلومات وتخزينها كميات هائلة من البيانات في معظم مجالات التطبيق ، مثل عالم الأعمال والمجتمع العلمي والطب والإدارة العامة. تمت الإشارة إلى مجموعة الأنشطة التي ينطوي عليها تحليل قواعد البيانات الكبيرة هذه ، عادةً بهدف استخراج المعرفة المفيدة لدعم اتخاذ القرار ، بطرق مختلفة ، مثل التنقيب عن البيانات واكتشاف المعرفة knowledge discovery والتعرف على الأنماط pattern recognition والتعلم الآلي machine learning.

على وجه الخصوص ، يشير مصطلح التنقيب عن البيانات إلى عملية استكشاف وتحليل مجموعة بيانات ، عادة ما تكون كبيرة الحجم ، من أجل العثور على أنماط منتظمة ، لاستخراج المعرفة ذات الصلة والحصول على قواعد متكررة ذات مغزى. يلعب التنقيب عن البيانات دورًا متزايدًا في كل من الدراسات النظرية والتطبيقات.

في هذا الفصل نرغب في وصف وتوصيف أنشطة التنقيب عن البيانات فيما يتعلق بأغراض التحقيق ومنهجيات التحليل. سيتم أيضًا مناقشة الخصائص ذات الصلة لبيانات الإدخال. أخيرًا ، سنصف عملية التنقيب عن البيانات وتوضيحها في مراحل متميزة.

5.1 تعريف التنقيب عن البيانات

تشكل أنشطة التنقيب عن البيانات عملية تكرارية تهدف إلى تحليل قواعد البيانات الكبيرة ، بغرض استخراج المعلومات والمعرفة التي قد تكون دقيقة ومفيدة للعاملين في مجال المعرفة المشاركين في صنع القرار وحل المشكلات.

كما هو موضح في القسم 5.3 ، فإن عملية التحليل تكرارية بطبيعتها نظرًا لوجود مراحل مميزة قد تتضمن التعليقات والمراجعات اللاحقة. عادةً ما تمثل هذه العملية نشاطًا تعاونيًا بين الخبراء في مجال التطبيق ومحللي البيانات ، الذين يستخدمون النماذج الرياضية للتعلم الاستقرائي. في الواقع ، تشير التجربة إلى أن دراسة التنقيب في البيانات تتطلب تدخلات متكررة من قبل المحللين عبر

مراحل التحقيق المختلفة ، وبالتالي لا يمكن أتمتتها بسهولة. من الضروري أيضًا أن تكون المعرفة المستخرجة دقيقة ، بمعنى أنه يجب تأكيدها بالبيانات وألا تؤدي إلى استنتاجات مضللة.

يشير مصطلح التنقيب عن البيانات إلى العملية الشاملة التي تتكون من جمع البيانات وتحليلها ، وتطوير نماذج التعلم الاستقرائي واعتماد القرارات العملية والإجراءات اللاحقة بناءً على المعرفة المكتسبة. مصطلح نظرية التعلم الرياضي mathematical learning theory مخصص لمجموعة متنوعة من النماذج والطرق الرياضية التي يمكن العثور عليها في صميم كل تحليل لتنقيب البيانات والتي تُستخدم لتوليد معرفة جديدة.

تعتمد عملية التنقيب عن البيانات على طرق التعلم الاستقرائي ، والغرض الرئيسي منها هو اشتقاق القواعد العامة بدءًا من مجموعة من الأمثلة المتاحة ، والتي تتكون من الملاحظات السابقة المسجلة في قاعدة بيانات واحدة أو أكثر. بعبارة أخرى ، الغرض من تحليل التنقيب في البيانات هو استخلاص بعض الاستنتاجات بدءًا من عينة من الملاحظات السابقة وتعميم هذه الاستنتاجات بالإشارة إلى جميع السكان ، بحيث تكون دقيقة قدر الإمكان. قد تتخذ النماذج والأنماط المحددة بهذه الطريقة أشكالًا مختلفة ، والتي سيتم وصفها في الفصول التالية ، مثل المعادلات الخطية ، ومجموعات القواعد في شكل if-then-else ، والمجموعات ، والمخططات ، والأشجار.

تعتمد خاصية أخرى للتنقيب في البيانات على إجراءات جمع الملاحظات السابقة وإدراجها في قاعدة بيانات. في الواقع ، عادةً ما يتم تخزين هذه السجلات لأغراض لا يتم توجيهها بشكل أساسي بواسطة تحليل التنقيب عن البيانات. على سبيل المثال ، سيتم تسجيل المعلومات الخاصة بالمشتريات من شركة البيع بالتجزئة ، أو حول استخدام كل رقم هاتف مخزن من قبل مزود الهاتف المحمول ، بشكل أساسي للأغراض الإدارية ، حتى إذا كان من الممكن استخدام البيانات لاحقًا لإجراء بعض التحليل المفيد لتنقيب البيانات. وبالتالي ، فإن إجراء جمع البيانات مستقل إلى حد كبير وغير مدرك لأهداف التنقيب عن البيانات ، بحيث يختلف اختلافًا جوهريًا عن أنشطة جمع البيانات التي يتم إجراؤها وفقًا لمخططات أخذ العينات المحددة مسبقًا ، النموذجية للإحصاءات الكلاسيكية. في هذا الصدد ، يمثل التنقيب عن البيانات شكلاً ثانويًا لتحليل البيانات.

يمكن تقسيم أنشطة التنقيب عن البيانات إلى مسارين رئيسيين للتحقيق ، وفقًا للغرض الرئيسي للتحليل: التفسير والتنبؤ.

- **الترجمة Interpretation.** الغرض من التفسير هو تحديد الأنماط المنتظمة في البيانات والتعبير عنها من خلال القواعد والمعايير التي يمكن فهمها بسهولة من قبل الخبراء في مجال التطبيق. يجب أن تكون القواعد التي تم إنشاؤها أصلية وغير تافهة من أجل زيادة مستوى المعرفة والفهم لنظام الاهتمام. على سبيل المثال ، بالنسبة لشركة في صناعة البيع بالتجزئة ، قد يكون من المفيد تجميع هؤلاء العملاء الذين حصلوا على بطاقات ولاء وفقًا لملف تعريف الشراء الخاص بهم. قد تكون الشرائح التي تم إنشاؤها بهذه الطريقة مفيدة في تحديد مجالات السوق الجديدة وتوجيه حملات التسويق المستقبلية.
- **التنبؤ Prediction.** الغرض من التنبؤ هو توقع القيمة التي سيفترضها المتغير العشوائي في المستقبل أو تقدير احتمالية الأحداث المستقبلية. على سبيل المثال ، قد يقوم مزود خدمة الهاتف المحمول بتطوير تحليل التنقيب في البيانات لتقدير احتمالية حدوث تقلب لصالح بعض المنافسين لعملائه. في سياق مختلف ، قد تتوقع شركة بيع بالتجزئة مبيعات منتج معين خلال الأسابيع التالية. في الواقع ، تستمد معظم تقنيات التنقيب عن البيانات تنبؤاتها من قيمة مجموعة من المتغيرات المرتبطة بالكيانات في قاعدة البيانات. على سبيل المثال ، قد يشير نموذج التنقيب عن البيانات إلى أن احتمال حدوث اضطراب في المستقبل للعميل يعتمد على ميزات مثل العمر ومدة العقد والنسبة المئوية للمكالمات لمشتركي موفري خدمات الهاتف الآخرين. ومع ذلك ، هناك نماذج السلاسل الزمنية الموضحة في الفصل 9 ، والتي تقوم بعمل تنبؤات تستند فقط إلى القيم السابقة لمتغير الاهتمام.

في بعض الأحيان ، قد يتضح أيضًا أن النموذج الذي تم تطويره لغرض التنبؤ يكون فعالًا في التفسير. في حالة أشجار التصنيف ذات قواعد تقسيم المحاور المتوازية ، والتي سيتم وصفها في الفصل 10 ، قد تكون النماذج التي تم إنشاؤها لأغراض تنبؤية مفيدة أيضًا في تحديد الظواهر التوضيحية المتكررة.

5.1.1 نماذج وأساليب تنقيب البيانات

هناك العديد من طرق التعلم المتاحة لأداء مهام التنقيب عن البيانات المختلفة. نشأ عدد من التقنيات في مجال علوم الكمبيوتر ، مثل أشجار التصنيف classification trees أو قواعد الارتباط association rules، ويشار إليها باسم التعلم الآلي machine learning أو اكتشاف المعرفة في قواعد البيانات knowledge discovery in databases. في معظم الحالات ، يميل النهج القائم على التجربة إلى السيادة ضمن هذه الفئة من التقنيات. تنتمي الأساليب الأخرى إلى الإحصائيات متعددة المتغيرات ، مثل الانحدار أو المصنفات البايزية ، وغالبًا ما تكون ذات طبيعة حدودية ولكنها تظهر على أسس نظرية أكثر.

تشمل التطورات الحديثة الأساليب الرياضية للتعلم ، مثل نظرية التعلم الإحصائي statistical learning theory، والتي تستند إلى أسس نظرية متينة وتضع نفسها على مفترق طرق مختلف التخصصات ، من بينها نظرية الاحتمالات ونظرية التحسين والإحصاء.

مثال 5.1 - الانحدار الخطي Linear regression. نماذج الانحدار الخطي ، الموصوفة في الفصل 8 ، هي واحدة من أشهر مناهج التعلم والتنبؤ في الإحصاء الكلاسيكي. في أبسط أشكاله ، يتم استخدام الانحدار الخطي لربط متغير استجابة تابع Y إلى متنبئ مستقل X من خلال انحدار خطي في الشكل $Y = aX + b$ ، حيث a و b معلمات يتم تحديدها باستخدام الملاحظات السابقة المتاحة. على سبيل المثال ، قد تمثل Y مبيعات منتج الاستهلاك الشامل خلال أسبوع و X إجمالي تكلفة الإعلان خلال نفس الأسبوع. فيما يتعلق بمراحل تطوير النموذج ، فإن اختيار الوظيفة الخطية يحدد نوع العلاقة بين المتنبئ ومتغير الاستجابة. مقياس التقييم المعقول هو مجموع الفروق التربيعية بين قيم Y التي لوحظت بالفعل في الماضي والقيم التي تنبأ بها النموذج الخطي. تحسب خوارزمية التحسين المناسبة قيمة المعلمتين a و b لتقليل مجموع الأخطاء التربيعية.

بغض النظر عن طريقة التعلم المحددة التي يرغب المرء في اعتمادها ، هناك خطوات متكررة أخرى في تطوير نموذج التنقيب عن البيانات ، كما هو موضح في المثال 5.1:

1. اختيار فئة من النماذج لاستخدامها في التعلم من الماضي و شكل محدد لتمثيل الأنماط في البيانات ؛

2. تعريف مقياس لتقييم فعالية ودقة النماذج التي يتم إنشاؤها.
3. تصميم خوارزمية حسابية لتوليد النماذج عن طريق تحسين مقياس التقييم.

في الفصول التالية سوف نقدم وصفًا لأكثر فئات طرق التنبؤ والتعرف على الأنماط شيوعًا. يمكن تطبيق بعض المنهجيات لأداء العديد من مهام التنقيب عن البيانات ، من بين تلك الموضحة في القسم 5.4. نظرًا لأنه من الممكن بشكل عام ربط كل فئة من النماذج بالنشاط السائد ، فسيتم ربط كل تقنية في الفصول التالية بالمهمة الأكثر ملاءمة.

5.1.2 تنقيب البيانات والإحصاءات الكلاسيكية و OLAP

تختلف مشاريع تنقيب البيانات في العديد من النواحي عن كل من الإحصاءات الكلاسيكية وتحليلات OLAP. تظهر هذه الاختلافات في الجدول 5.1 ، مع الإشارة إلى مثال.

Table 5.1 Differences between OLAP, statistics and data mining

OLAP	statistics	data mining
extraction of details and aggregate totals from data	verification of hypotheses formulated by analysts	identification of patterns and recurrences in data
information	validation	knowledge
distribution of incomes of home loan applicants	analysis of variance of incomes of home loan applicants	characterization of home loan applicants and prediction of future applicants

الجدول 5.1 الاختلافات بين OLAP والإحصاءات وتنقيب البيانات

يتكون الاختلاف الرئيسي من التوجيه النشط الذي توفره نماذج التعلم الاستقرائي ، مقارنة بالطبيعة السلبية للتقنيات الإحصائية و OLAP. في الواقع ، في التحليلات الإحصائية ، يصيغ صانعو القرار فرضية يجب تأكيدها بعد ذلك على أساس عينة من الأدلة. وبالمثل ، في تحليلات OLAP ، يعبر عمال المعرفة عن بعض الحدس الذي يستندون إليه في معايير الاستخراج وإعداد التقارير والتصور. كلا الطريقتين - من ناحية تقنيات التحقق من الصحة الإحصائية ومن ناحية أخرى أدوات المعلومات للتنقل عبر مكعبات البيانات - توفر فقط عناصر لتأكيد أو دحض الفرضيات التي صاغها صانع القرار ، وفقًا لتدفق التحليل من أعلى إلى أسفل. على العكس من ذلك ، فإن نماذج التعلم ، التي تمثل جوهر مشاريع التنقيب عن البيانات ، قادرة على لعب دور نشط من خلال توليد تنبؤات وتفسيرات تمثل في الواقع معرفة جديدة متاحة للمستخدمين.

تدفق التحليل في الحالة الأخيرة له هيكل من أسفل إلى أعلى. على وجه الخصوص ، عند مواجهة كميات كبيرة من البيانات ، يصبح استخدام النماذج القادرة على لعب دور فعال عامل نجاح حاسم ، حيث يصعب على العاملين في مجال المعرفة صياغة فرضيات مسبقة ذات مغزى وراسخة.

5.1.3 تطبيقات التنقيب عن البيانات

يمكن تطبيق منهجيات التنقيب عن البيانات على مجموعة متنوعة من المجالات ، من التحكم في عمليات التسويق والتصنيع إلى دراسة عوامل الخطر في التشخيص الطبي ، ومن تقييم فعالية الأدوية الجديدة إلى اكتشاف الاحتيال.

- **التسويق العلائقي Relational marketing.** ساهمت تطبيقات التنقيب عن البيانات في مجال التسويق العلائقي ، الموصوف في الفصل 13 ، بشكل كبير في زيادة شعبية هذه المنهجيات. بعض التطبيقات ذات الصلة في التسويق العلائقي هي:
  - تحديد قطاعات العملاء التي من المرجح أن تستجيب لحملات التسويق المستهدفة ، مثل البيع المتبادل والبيع الإضافي. •
  - تحديد شرائح العملاء المستهدفة لحملات الاحتفاظ.
  - التنبؤ بمعدل الاستجابات الإيجابية للحملات التسويقية.
  - تفسير وفهم السلوك الشرائي للعميل.
  - تحليل المنتجات المشتراة بشكل مشترك من قبل العملاء ، والمعروفة باسم تحليل سلة التسوق Market basket analysis.
- **الكشف عن الغش Fraud detection.** يعد اكتشاف الاحتيال مجالًا آخر ذي صلة بتطبيق التنقيب عن البيانات. قد يؤثر الاحتيال على صناعات مختلفة مثل الاتصالات الهاتفية والتأمين (الادعاءات الكاذبة) والخدمات المصرفية (الاستخدام غير القانوني لبطاقات الائتمان والشيكات المصرفية ؛ المعاملات النقدية غير القانونية).
- **تقييم الخطر Risk evaluation.** الغرض من تحليل المخاطر هو تقدير المخاطر المرتبطة بالقرارات المستقبلية ، والتي غالبًا ما تتخذ شكلاً ثنائي التفرع. على سبيل المثال ، باستخدام الملاحظات السابقة المتاحة ، قد يقوم البنك بتطوير نموذج تنبؤي لتحديد ما

إذا كان من المناسب منح قرض نقدي أو قرض سكني ، بناءً على خصائص مقدم الطلب.

- **تنقيب النصوص Text mining**. يمكن تطبيق التنقيب عن البيانات على أنواع مختلفة من النصوص ، والتي تمثل بيانات غير منظمة ، من أجل تصنيف المقالات والكتب والوثائق ورسائل البريد الإلكتروني وصفحات الويب. ومن الأمثلة على ذلك محركات البحث على الويب أو التصنيف التلقائي للبيانات الصحفية لأغراض التخزين. تتضمن تطبيقات التنقيب عن النصوص الأخرى إنشاء عوامل تصفية لرسائل البريد الإلكتروني ومجموعات الأخبار.
- **التعرف على الصور Image recognition**. تعتبر معالجة الصور الرقمية وتصنيفها ، سواء كانت ثابتة أو ديناميكية ، موضوعًا مثيرًا لكل من اهتماماتها النظرية والعدد الكبير من التطبيقات التي تقدمها. من المفيد التعرف على الحروف المكتوبة ومقارنة وتحديد الوجوه البشرية وتطبيق مرشحات التصحيح على معدات التصوير واكتشاف السلوكيات المشبوهة من خلال كاميرات فيديو المراقبة.
- **التنقيب على الويب Web mining**. تهدف تطبيقات التنقيب على الويب ، التي سيتم النظر فيها بإيجاز في القسم 13.1.9 ، إلى تحليل ما يسمى بـ clickstreams - تسلسل الصفحات التي تمت زيارتها والخيارات التي يقوم بها متصفحي الويب. قد تكون مفيدة لتحليل مواقع التجارة الإلكترونية ، في تقديم صفحات مرنة ومخصصة لمتصفحي ، في التخزين المؤقت للصفحات الأكثر شعبية أو في تقييم فعالية دورة تدريبية للتعلم الإلكتروني.
- **تشخيص طبي Medical diagnosis**. نماذج التعلم هي أداة لا تقدر بثمن في المجال الطبي للكشف المبكر عن الأمراض باستخدام نتائج الاختبارات السريرية. يعد تحليل الصور لأغراض التشخيص مجالًا آخر من مجالات التحقيق يزدهر حاليًا.

5.2 تمثيل بيانات الإدخال

في معظم الحالات ، يأخذ الإدخال إلى تحليل تنقيب البيانات شكل جدول ثنائي الأبعاد ، يسمى مجموعة البيانات ، بغض النظر عن المنطق الفعلي وتمثيل المواد المعتمد لتخزين المعلومات

في الملفات وقواعد البيانات ومستودعات البيانات والبيانات مارت تستخدم كمصادر البيانات. تتوافق الصفوف الموجودة في مجموعة البيانات مع الملاحظات observations المسجلة في الماضي وتسمى أيضًا أمثلة examples, أو حالات cases, أو مثيلات instances أو سجلات records. تمثل الأعمدة المعلومات المتاحة لكل ملاحظة ويطلق عليها سمات attributes أو متغيرات variables أو خصائص characteristics أو ميزات features.

يمكن تصنيف السمات الموجودة في مجموعة البيانات على أنها فئوية categorical أو رقمية numerical، اعتمادًا على نوع القيم التي تأخذها.

- **الخصائص الفئوية categorical** . تفترض السمات الفئوية عددًا محدودًا من القيم المميزة ، في معظم الحالات يقتصر على أقل من مائة ، تمثل خاصية نوعية لكيان تشير إليه. أمثلة على السمات الفئوية هي مقاطعة إقامة الفرد (والتي تأخذ كقيم سلسلة من الأسماء ، والتي بدورها قد يتم تمثيلها بأعداد صحيحة) أو ما إذا كان العميل قد تخلى عن مزود الخدمة (المعبر عنه بالقيمة 1) أو ظل مخلصًا إليها (معبرًا عنها بالقيمة 0). لا يمكن تطبيق العمليات الحسابية على السمات الفئوية حتى عندما يتم التعبير عن ترميز قيمها بأرقام صحيحة.
- **الخصائص العددية numerical**. تفترض السمات العددية عددًا محدودًا أو لا نهائيًا من القيم وتفسح المجال لعمليات الطرح أو القسمة. على سبيل المثال ، يمثل مقدار المكالمات الهاتفية الصادرة خلال شهر لعميل عام متغيرًا رقميًا. فيما يتعلق بعميلين A و B يجريان مكالمات هاتفية في أسبوع مقابل 27 و 36 على التوالي ، فمن المنطقي أن ندعي أن الفرق بين المبالغ التي أنفقها العميلان يساوي $9 وأن A قد أنفق ثلاثة أرباع المبلغ الذي أنفقته B.

في بعض الأحيان ، يمكن أن يكون تصنيف السمات الأكثر دقة مفيدًا.

- **العد Counts**. الأعداد هي سمات فئوية فيما يتعلق بخاصية معينة يمكن أن تكون صحيحة أو خاطئة. لذلك يمكن تمثيل هذه السمات باستخدام المتغيرات المنطقية {true ،false} أو المتغيرات الثنائية {0،1}. على سبيل المثال ، قد يكون عملاء البنك حاملين أو لا يحملون بطاقة ائتمان صادرة عن البنك.

- **اسمي Nominal:** السمات الاسمية هي سمات فئوية بدون ترتيب طبيعي ، مثل مقاطعة الإقامة.
- **ترتيبي Ordinal**. السمات الترتيبية ، مثل مستوى التعليم ، هي سمات فئوية تصلح للترتيب الطبيعي ولكن ليس من المنطقي حساب الفروق أو النسب بين القيم.
- **منفصله Discrete**. السمات المنفصلة هي سمات عددية تفترض عددًا محددًا أو عددًا لا نهائيًا من القيم القابلة للعد.

بمعنى آخر ، المجموعة قابلة للعد إذا كان هناك انحراف بين تلك المجموعة و N. توجد مجموعات ، مثل المجموعة R للأرقام الحقيقية ، والتي لا نهائية وغير قابلة للعد ، وبالتالي تسمى غير قابلة للعد.

- **مستمر Continuous**. السمات المستمرة هي سمات عددية تفترض عددًا لا نهائيًا من القيم.

لتمثيل مجموعة بيانات عامة D ، سوف نشير بواسطة m عدد الملاحظات ، أو الصفوف ، في الجدول ثنائي الأبعاد الذي يحتوي على البيانات و n عدد السمات ، أو الأعمدة. علاوة على ذلك ، سوف نشير إلى

$$\mathbf{X} = [x_{ij}], \quad i \in \mathcal{M} = \{1, 2, \ldots, m\}, \quad j \in \mathcal{N} = \{1, 2, \ldots, n\}, \tag{5.1}$$

مصفوفة الأبعاد m × n التي تتوافق مع الإدخالات في مجموعة البيانات D. سيكون:

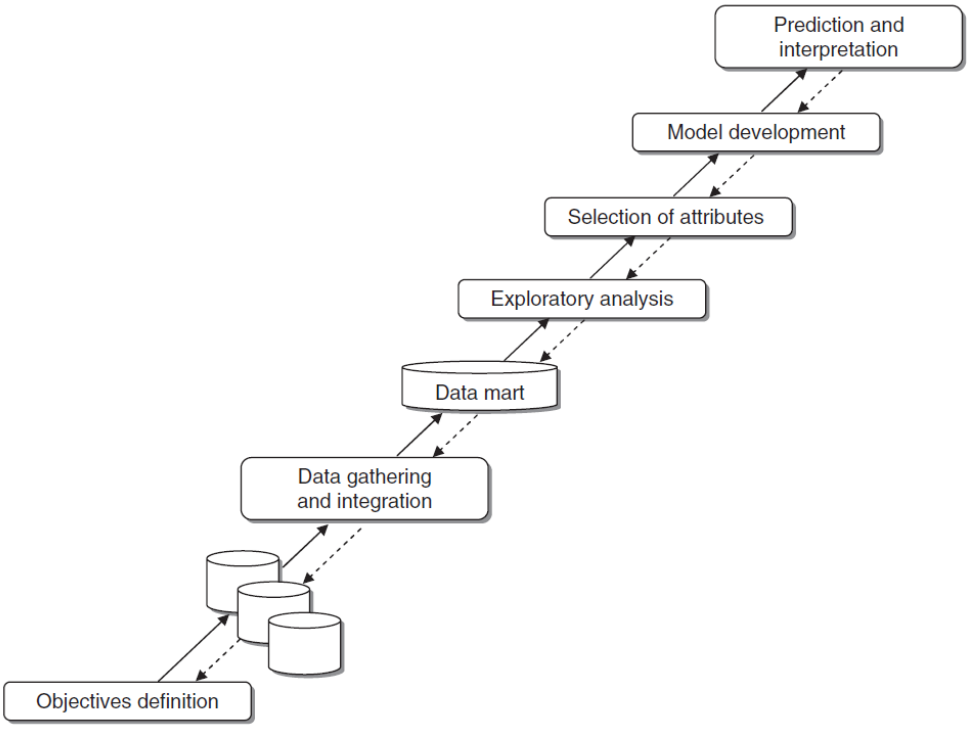
$$\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{in}) \tag{5.2}$$

$$\mathbf{a}_j = (x_{1j}, x_{2j}, \ldots, x_{mj}) \tag{5.3}$$

حيث أن x_i تمثل متجه الصف المحتوي على n من الأبعاد والمرتبط بالقيد i في مجموعة البيانات D و a_j يمثل متجه العمود المحتوي على m من الأبعاد والذي يمثل الصفة بالتسلسل j في مجموعة البيانات D.

5.3 عملية التنقيب عن البيانات

يشير تعريف تنقيب البيانات الوارد في بداية القسم 5.1 إلى عملية تكرارية ، تلعب خلالها نماذج وتقنيات التعلم دورًا رئيسيًا ، وإن كان غير شامل. يوضح الشكل 5.1 المراحل الرئيسية لعملية التنقيب عن البيانات العامة.



الشكل 5.1 عملية التنقيب عن البيانات

- **تعريف الأهداف.** يتم إجراء تحليلات التنقيب عن البيانات في مجالات تطبيق محددة وتهدف إلى تزويد صانعي القرار بمعرفة مفيدة. ونتيجة لذلك ، فإن الحدس والكفاءة مطلوبان من قبل خبراء المجال من أجل صياغة أهداف تحقيق معقولة ومحددة جيدًا. إذا لم يتم تحديد المشكلة المطروحة بشكل كافٍ وقيدها ، فقد يتعرض المرء لخطر إحباط أي جهد مستقبلي يُبذل أثناء أنشطة التنقيب عن البيانات. سيستفيد تحديد الأهداف من التعاون الوثيق بين الخبراء في مجال التطبيق ومحللي التنقيب عن البيانات. بالإشارة إلى المثال 5.2 ، من الممكن تحديد المشكلة وأهداف التحقيق كتحليل البيانات السابقة وتحديد نموذج للتعبير عن ميل العملاء لترك الخدمة بناءً على خصائصهم ، من أجل فهم أسباب عدم الولاء والتنبؤ باحتمالية حدوث اضطراب.

مثال 5.2 - الاحتفاظ في صناعة الهاتف المحمول.

يوضح الجدول 5.2 البنية ثنائية الأبعاد لبيانات الإدخال من مثال لتحليل ولاء العملاء. افترض أن شركة الهاتف المحمول تقوم بإجراء تحليل التنقيب عن البيانات مع أهداف التنبؤ والتفسير. من ناحية أخرى ، ترغب الشركة في تقييم احتمالية حدوث اضطراب في المستقبل من قبل كل عميل ، من أجل استهداف إجراءات التسويق لأغراض الاحتفاظ. من ناحية أخرى ، فإن القصد هو فهم أسباب تضاؤل العملاء ، بهدف تحسين مستوى الخدمة وتقليل الاضطراب في المستقبل. يحتوي الجدول 5.2 على 23 ملاحظة و 12 سمة ، يشار إلى معناها في الجدول 5.3.

الجدول 5.2 مثال على بيانات الإدخال لنموذج التنقيب عن البيانات

area	numin	timein	numout	Pothers	Pmob	Pland	numsms	numserv	numcall	diropt	churner
3	32	8093	45	0.14	0.75	0.12	18	1	0	0	0
3	277	157842	450	0.26	0.35	0.38	9	3	0	1	0
1	17	15023	20	0.37	0.23	0.40	1	1	0	0	0
1	46	22459	69	0.10	0.39	0.51	33	1	0	0	0
1	19	8640	9	0.00	0.00	1.00	0	0	0	0	0
2	17	7652	66	0.16	0.42	0.43	1	3	0	1	0
3	47	17768	11	0.45	0.00	0.55	0	0	0	0	0
3	19	9492	42	0.18	0.34	0.48	3	1	0	0	1
1	1	84	9	0.09	0.54	0.37	0	0	0	0	1
2	119	87605	126	0.84	0.02	0.14	12	1	0	0	0
4	24	6902	47	0.25	0.26	0.48	4	1	0	0	0
1	32	28072	43	0.28	0.66	0.06	0	1	0	0	0
3	103	112120	24	0.61	0.28	0.11	24	2	0	0	0
3	45	21921	94	0.34	0.47	0.19	45	2	0	1	0
1	8	25117	89	0.02	0.89	0.09	189	1	3	0	0
3	4	945	16	0.00	0.00	1.00	0	0	0	0	1
2	83	44263	83	0.00	0.00	0.67	0	0	0	0	1
2	22	15979	59	0.05	0.53	0.41	5	2	0	1	1
2	0	0	57	0.00	1.00	0.00	15	1	1	0	1
4	162	114108	273	0.18	0.15	0.41	2	3	0	1	1
4	21	4141	70	0.14	0.58	0.28	0	1	0	1	1
4	33	10066	45	0.12	0.21	0.67	0	0	0	0	1
4	5	965	40	0.41	0.27	0.32	64	1	0	0	1

تمثل السمة الـ 11 الأولى متغيرات توضيحية ، بينما تمثل السمة الأخيرة المتغير المستهدف ، معبرة عن فئة كل سجل فيما يتعلق بأهداف تحليل تنقيب البيانات. يعطي المتغير التوضيحي الأول معلومات ديموغرافية شخصية بينما يشير الباقي إلى استخدام الخدمة. القيم المرصودة تتعلق بالفترة الزمنية للمؤشر t - 2 للسمات التفسيرية ، بينما بالنسبة للمتغير المستهدف فإنها تشير إلى الفترة الزمنية t. الفرق في تحديد الوقت مطلوب من أجل استخدام النموذج للأغراض التنبؤية. في الواقع ، من الضروري التنبؤ خلال الفترة الحالية بمن سيغادر العميل خلال فترتين ، بناءً على المعلومات المتاحة ، من أجل تطوير إجراءات الاحتفاظ الفعالة في الوقت المناسب ؛ راجع أيضًا مناقشة وقت الاستجابة الزمني للنماذج التنبؤية في القسم 13.1.4.

بالطبع ، في التطبيق الحقيقي ، يكون عدد السمات المتاحة أعلى بكثير ، بترتيب المئات أو الآلاف ، وعدد الصفوف التي تمثل customers أكبر بكثير ، من مئات الآلاف أو ملايين السجلات. الغرض هنا هو الحصول على نموذج استقرائي قادر على التعلم من الملاحظات السابقة المتاحة وتحديد علاقة معقولة بين المتغير الهدف والسمات التفسيرية.

بمجرد إنشاء النموذج استنادًا إلى السجلات السابقة ، من الممكن استخدامه للتنبؤ بالفئة المستهدفة من السجلات الجديدة أو لفهم الخصائص المشتركة للعملاء الذين يتصرفون وفقًا للولاء مقارنة بأولئك الذين يظلون مخلصين.

الجدول 5.3 معنى السمات الواردة في الجدول 5.2

السمة	معناها
numin	عدد المكالمات المستلمة في الفترة t - 2
timein	المدة بالثواني للمكالمات المستلمة في الفترة t - 2
numout	عدد المكالمات التي تم إجراؤها في الفترة t - 2
Pother	النسبة المئوية للمكالمات التي تم إجراؤها لشركات الهاتف المحمول الأخرى في الفترة t - 2
Pmob	النسبة المئوية للمكالمات التي تمت على نفس شركة الهاتف المحمول في الفترة t - 2
Pland	النسبة المئوية للمكالمات التي تم إجراؤها على أرقام الأرض في الفترة t - 2
numsms	عدد الرسائل المرسلة في الفترة t - 2
numserv	عدد المكالمات التي تم إجراؤها على الخدمات الخاصة في الفترة t - 2
numcall	عدد المكالمات التي تم إجراؤها على مركز الاتصال في الفترة t - 2

diropt	متغير ثنائي يشير إلى ما إذا كان العميل المقابل للسجل قد اشترك في خطة أسعار خاصة للمكالمات التي تم إجراؤها على أرقام محددة
churner	متغير ثنائي يشير إلى ما إذا كان العميل المقابل للسجل قد ترك الخدمة في الفترة t

- **جمع البيانات وتكاملها Data Gathering and integration :** بمجرد تحديد أهداف التحقيق ، يبدأ جمع البيانات. قد تأتي البيانات من مصادر مختلفة وبالتالي قد تتطلب التكامل. قد تكون مصادر البيانات داخلية أو خارجية أو مزيجًا من الاثنين. قد يتم اقتراح تكامل مصادر البيانات المتميزة من خلال الحاجة إلى إثراء البيانات بأبعاد وصفية جديدة ، مثل متغيرات التسويق الجغرافي ، أو مع قوائم بأسماء العملاء المحتملين ، والتي يطلق عليها الآفاق ، والتي لم تكن موجودة بعد في نظام معلومات الشركة. في بعض الحالات ، تكون مصادر البيانات منظمة بالفعل في مستودعات البيانات وخزانات البيانات لتحليلات OLAP وبشكل أكثر عمومية لأنشطة دعم القرار. هذه مواقف مواتية حيث يكفي تحديد السمات التي تعتبر ذات صلة لغرض تحليل التنقيب في البيانات. ومع ذلك ، هناك خطر يتمثل في أنه من أجل الحد من استيعاب الذاكرة ، تم تجميع المعلومات المخزنة في مستودع البيانات وتوحيدها إلى الحد الذي يجعل أي تحليل لاحق عديم الفائدة. على سبيل المثال ، إذا كانت الشركة في صناعة البيع بالتجزئة تخزن لكل عميل المبلغ الإجمالي لكل إيصال ، دون تتبع كل عنصر تم شراؤه فرديًا ، فقد يتم اختراق تحليل تنقيب البيانات المستقبلي الذي يهدف إلى التحقيق في سلوك الشراء الفعلي.

في حالات أخرى ، تحتوي البيانات الأصلية على تنسيق غير متجانس بدون بنية محددة مسبقًا. في هذه الحالة ، تصبح عملية جمع البيانات والتكامل أكثر صعوبة وبالتالي أكثر عرضة للأخطاء. بغض النظر عن الهيكل الأصلي ، فإن مجموعات البيانات المدخلة لتحليلات التنقيب عن البيانات تأخذ دائمًا شكل جداول ثنائية الأبعاد ، كما هو مذكور أعلاه. على عكس العديد من إجراءات أخذ العينات القياسية للإحصاءات الكلاسيكية ، تمثل مجموعات البيانات لتنقيب

البيانات عينات مستخرجة وفقًا لتوزيع غير معروف ، مع عدم قدرة المحللين على التأثير على عملية جمع البيانات. الفصل 6 سيناقش قضايا إعداد البيانات بمزيد من التفصيل.

- **التحليل الاستكشافي Exploratory analysis.** في المرحلة الثالثة من عملية التنقيب عن البيانات ، يتم إجراء تحليل أولي للبيانات بغرض التعرف على المعلومات المتاحة وتنفيذ تنقية البيانات. عادة ، تتم معالجة البيانات المخزنة في مستودع البيانات في وقت التحميل بطريقة تزيل أي تناقضات نحوية. على سبيل المثال ، يتم اكتشاف وتصحيح تواريخ الميلاد التي تقع خارج النطاقات المسموح بها ورسوم المبيعات السلبية. في عملية التنقيب عن البيانات ، يحدث تنظيف البيانات على مستوى دلالي.

في البداية ، يتم دراسة توزيع القيم لكل سمة ، باستخدام الرسوم البيانية للسمات الفئوية وإحصاءات الملخص الأساسية للمتغيرات العددية. بهذه الطريقة ، يتم أيضًا تمييز أي قيم غير طبيعية (القيم المتطرفة) والقيم المفقودة. تتم دراستها من قبل خبراء في مجال التطبيق والذين قد يفكرون في استبعاد السجلات المقابلة من التحقيق.

الفصل 7 سيناقش التقنيات المستخدمة لتطوير تحليل البيانات الاستكشافية.

- **تحديد السمة Attribute Selection.** في المرحلة اللاحقة ، يتم تقييم أهمية السمات المختلفة فيما يتعلق بأهداف التحليل. تتم إزالة السمات التي ثبت أنها ذات فائدة قليلة ، وذلك لتنظيف المعلومات غير ذات الصلة من مجموعة البيانات. علاوة على ذلك ، يتم تضمين سمات جديدة تم الحصول عليها من المتغيرات الأصلية من خلال عمليات التحويل المناسبة في مجموعة البيانات. على سبيل المثال ، من المفيد في معظم الحالات تقديم سمات جديدة تعكس الاتجاهات المتأصلة في البيانات من خلال حساب النسب والاختلافات بين المتغيرات الأصلية. يعد التحليل الاستكشافي واختيار السمات مراحل حاسمة وغالبًا ما تكون صعبة في عملية تنقيب البيانات وقد تؤثر إلى حد كبير على مستوى نجاح المراحل اللاحقة. قد تكون الطرق الموصوفة في الفصلين 6 و 7 مفيدة في تحويل واختيار السمات.
- **تطوير النموذج والتحقق من صحته Model development and validation.** بمجرد تجميع مجموعة بيانات عالية الجودة وربما إثرائها بسمات محددة حديثًا ، يمكن تطوير التعرف على الأنماط والنماذج التنبؤية. عادة ما يتم تدريب النماذج باستخدام عينة من

السجلات المستخرجة من مجموعة البيانات الأصلية. بعد ذلك ، يمكن تقييم الدقة التنبؤية لكل نموذج تم إنشاؤه باستخدام باقي البيانات. بتعبير أدق ، يتم تقسيم مجموعة البيانات المتاحة إلى مجموعتين فرعيتين. الأول يشكل مجموعة التدريب ويستخدم لتحديد نموذج تعليمي معين ضمن فئة النماذج المختارة. عادةً ما يتم اختيار حجم عينة مجموعة التدريب ليكون صغيرًا نسبيًا ، على الرغم من أهميته من وجهة نظر إحصائية - على سبيل المثال ، بضعة آلاف من الملاحظات. المجموعة الفرعية الثانية هي مجموعة الاختبار وتستخدم لتقييم دقة النماذج البديلة التي تم إنشاؤها أثناء مرحلة التدريب ، من أجل تحديد أفضل نموذج للتنبؤات المستقبلية الفعلية.

ستتم مناقشة الفئات الأكثر شيوعًا لنماذج التعلم بالتفصيل في الفصول التالية.

- **التنبؤ والتفسير Prediction and interpretation.** عند الانتهاء من عملية تنقيب البيانات ، يجب تنفيذ النموذج المختار من بين تلك التي تم إنشاؤها خلال مرحلة التطوير واستخدامه لتحقيق الأهداف التي تم تحديدها في الأصل. علاوة على ذلك ، يجب دمجها في الإجراءات الداعمة لعمليات صنع القرار حتى يتمكن العاملون في مجال المعرفة من استخدامها لرسم التنبؤات واكتساب معرفة أكثر عمقًا بالظاهرة محل الاهتمام.

تتضمن عملية تنقيب البيانات دورات التغذية الراجعة ، التي تمثلها الأسهم المنقطة في الشكل 5.1 ، والتي قد تشير إلى العودة إلى مرحلة سابقة ، اعتمادًا على نتيجة المراحل اللاحقة.

أخيرًا ، يجب أن نؤكد على أهمية إشراك وتفاعل العديد من الأدوار المهنية من أجل تحقيق عملية استخراج بيانات فعالة:

- خبير في مجال التطبيق ، من المتوقع أن يحدد الأهداف الأصلية للتحليل ، لتوفير الفهم المناسب أثناء أنشطة التنقيب عن البيانات اللاحقة والمساهمة في اختيار النموذج الأكثر فعالية ودقة ؛
- خبير في أنظمة معلومات الشركة ، من المتوقع أن يشرف على الوصول إلى مصادر المعلومات.
- خبير في النظرية الرياضية للتعلم والإحصاء ، لتحليل البيانات الاستكشافية وتوليد النماذج التنبؤية.

يوضح الشكل 5.2 الكفاءات والقواعد الداخلة في الأنشطة المختلفة لكل جهة فاعلة في عملية التنقيب عن البيانات.

5.4 منهجيات التحليل

يمكن تقسيم أنشطة التنقيب عن البيانات إلى فئات رئيسية قليلة ، بناءً على مهام وأهداف التحليل. اعتمادًا على الوجود المحتمل لمتغير مستهدف ، يمكن للمرء أن يرسم أول تمييز أساسي بين عمليات التعلم الخاضعة للإشراف وغير الخاضعة للإشراف.

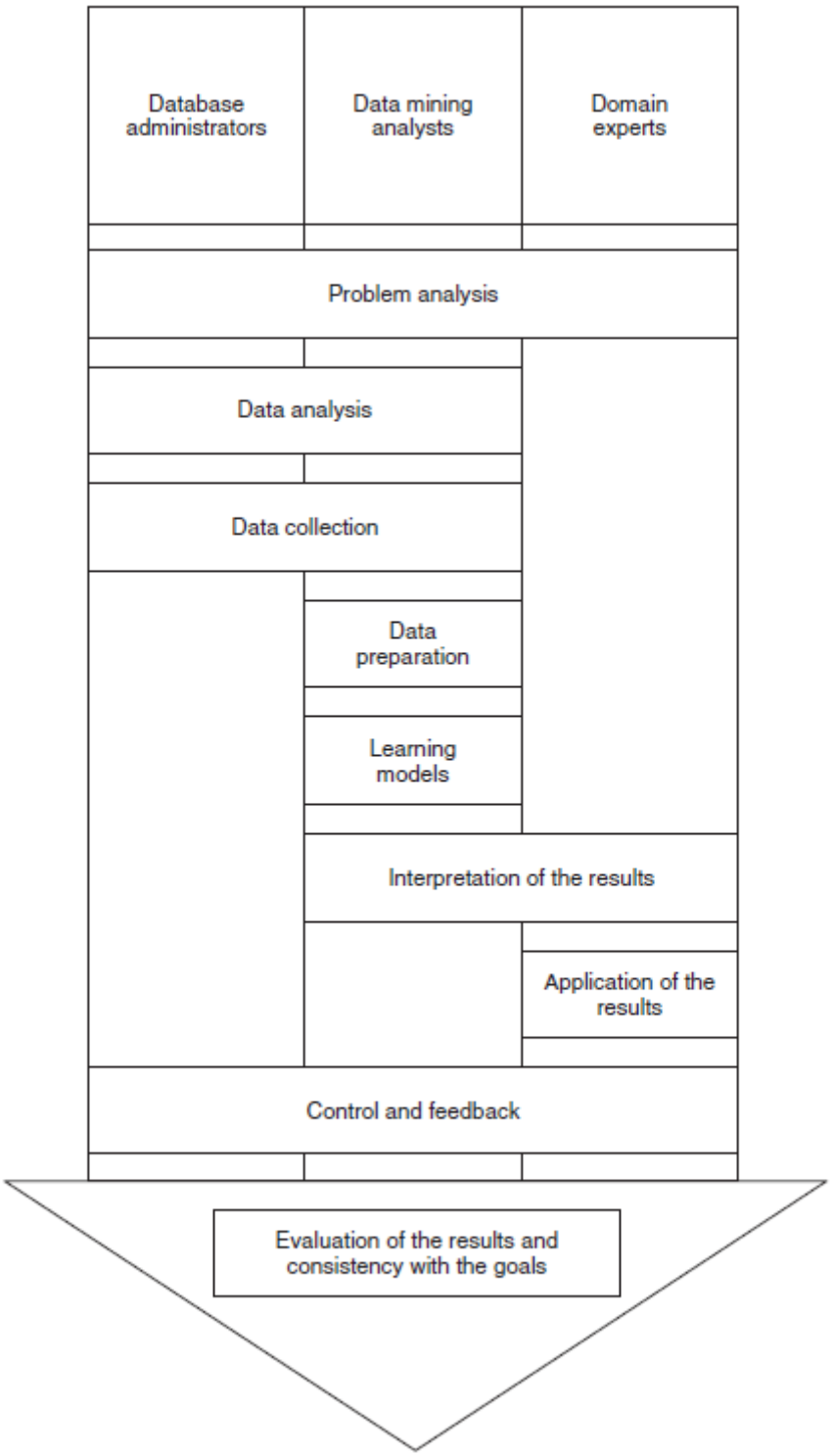
- **التعلم تحت الإشراف Supervised learning**. في تحليل التعلم الخاضع للإشراف (أو المباشر) ، تمثل السمة الهدف إما الفئة التي ينتمي إليها كل سجل ، كما هو موضح في المثال 5.2 حول الولاء في صناعة الهاتف المحمول ، أو تعبر عن كمية قابلة للقياس ، مثل القيمة الإجمالية للمكالمات التي سيتم وضعها من قبل العميل في فترة مقبلة. كمثال ثانٍ للمنظور الخاضع للإشراف ، ضع في اعتبارك شركة إدارة استثمار ترغب في التنبؤ بالميزانية العمومية لعملائها بناءً على خصائصهم الديموغرافية ومعاملات الاستثمار السابقة. وبالتالي فإن عمليات التعلم الخاضعة للإشراف موجهة نحو التنبؤ والتفسير فيما يتعلق بالسمة المستهدفة.
- **تعليم غير مشرف عليه Unsupervised learning**. لا تسترشد تحليلات التعلم غير الخاضعة للإشراف (أو غير المباشر) بالسمة المستهدفة. لذلك ، تهدف مهام التنقيب عن البيانات في هذه الحالة إلى اكتشاف الأنماط المتكررة والصلات في مجموعة البيانات. كمثال ، ضع في اعتبارك شركة إدارة استثمار ترغب في تحديد مجموعات العملاء الذين يظهرون سلوكًا استثماريًا متجانسًا ، استنادًا إلى بيانات عن المعاملات السابقة. في معظم تحليلات التعلم غير الخاضعة للإشراف ، يهتم المرء بتحديد مجموعات السجلات المتشابهة داخل كل مجموعة ومختلفة عن أعضاء المجموعات الأخرى.

بمزيد من التمييز ، يمكن تحديد سبع مهام أساسية للتنقيب عن البيانات:

1. التوصيف والتمييز characterization and discrimination
2. تصنيف classification
3. الأنحدار regression

4. تحليل السلاسل الزمنية time series analysis
5. قواعد الترابط Association rules
6. العنقدة Clustering
7. الوصف والتصور Description and visualization

تتوافق المهام الأربع الأولى مع تحليلات التنقيب عن البيانات الخاضعة للإشراف ، نظراً لوجود متغير هدف محدد يجب شرحه بناءً على السمات المتاحة أو طوال تطوره بمرور الوقت. تمثل المهام الثلاث المتبقية تحليلات غير خاضعة للإشراف الغرض منها تطوير نماذج قادرة على التعبير عن العلاقة المتبادلة بين السمات المتاحة.



الشكل 5.2 الجهات الفاعلة والأدوار في عملية التنقيب عن البيانات

- **التوصيف والتمييز characterization and discrimination**. في حالة وجود سمة هدف فئوية ، قبل البدء في تطوير نموذج تصنيف ، يكون من المفيد غالبًا إجراء تحليل

استكشافي غرضه مزدوج. من ناحية أخرى ، الهدف هو تحقيق توصيف من خلال مقارنة توزيع قيم السمات للسجلات التي تنتمي إلى نفس الفئة. من ناحية أخرى ، الغرض هو اكتشاف الاختلاف ، من خلال المقارنة بين توزيع قيم السمات لسجلات فئة معينة وسجلات فئة مختلفة ، أو بين سجلات فئة معينة وجميع السجلات المتبقية. يتم إجراء مهمة التنقيب عن البيانات هذه في المقام الأول عن طريق تحليل البيانات الاستكشافية ، وبالتالي فهي تستند إلى استفسارات وحسابات لا تتطلب تطوير نماذج تعليمية محددة. وعادة ما يتم تقديم المعلومات التي يتم الحصول عليها إلى المستخدمين في شكل رسوم بيانية وأنواع أخرى من الرسوم البيانية ، كما هو موضح في الفصل 7. ومع ذلك ، فإن قيمة المعلومات المتولدة ملحوظة وقد توجه في كثير من الأحيان المرحلة اللاحقة من اختيار السمة.

- **تصنيف classification**. في مشكلة التصنيف ، تتوفر مجموعة من الملاحظات ، وعادة ما يتم تمثيلها من خلال سجلات مجموعة البيانات ، والتي يُعرف صنفها المستهدف.

قد تتوافق الملاحظات ، على سبيل المثال ، مع عملاء الهاتف المحمول وقد تشير الفئة الثنائية إلى ما إذا كان عميل معين لا يزال نشطًا أو قد تحرك. يتم وصف كل ملاحظة من خلال عدد معين من السمات التي تُعرف قيمتها ؛ في المثال السابق ، قد تتوافق السمات مع العمر وأقدمية العميل وحركة الهاتف الصادرة المميزة بالوجهة.

لذلك يمكن لخوارزمية التصنيف استخدام الملاحظات المتاحة المتعلقة بالماضي من أجل تحديد نموذج يمكنه التنبؤ بالفئة المستهدفة للملاحظات المستقبلية التي تُعرف قيم سماتها. من الجدير بالذكر أن السمة الهدف ، التي سيتم التنبؤ بقيمتها ، تعتبر قاطعة في مشاكل التصنيف ، وبالتالي فهي تأخذ عددًا محدودًا من القيم وعادةً ما يكون صغيرًا إلى حد ما.

في معظم التطبيقات يتم تمثيل الهدف حتى بواسطة متغير ثنائي. تحدد الطبيعة الفئوية للهدف التمييز بين التصنيف والانحدار.

- **الأنحدار regression**. على عكس التصنيف ، المخصص للأهداف المنفصلة ، يتم استخدام الانحدار عندما يأخذ المتغير الهدف قيمًا مستمرة. استنادًا إلى السمات التوضيحية المتاحة ، يتمثل الهدف في التنبؤ بقيمة المتغير المستهدف لكل ملاحظة. إذا رغب المرء في التنبؤ بمبيعات منتج ما بناءً على الحملات الترويجية المركبة وسعر البيع ،

فقد يأخذ المتغير المستهدف عددًا كبيرًا جدًا من القيم المنفصلة ويمكن التعامل معه على أنه متغير مستمر. يمكن تحويل مشكلة التصنيف إلى مشكلة انحدار والعكس صحيح. لمعرفة ذلك ، قد تواجه شركة الهاتف المحمول المهتمة بتصنيف العملاء بناءً على ولائهم مشكلة انحدار من خلال التنبؤ باحتمالية بقاء كل عميل مخلصًا.

- **تحليل السلاسل الزمنية time series analysis** : تتطور السمة الهدف أحيانًا بمرور الوقت وبالتالي فهي مرتبطة بالفترات المجاورة على محور الوقت. في هذه الحالة ، يُقال إن تسلسل قيم المتغير المستهدف يمثل سلسلة زمنية. على سبيل المثال ، تمثل المبيعات الأسبوعية لمنتج معين التي تمت ملاحظتها على مدار عامين سلسلة زمنية تحتوي على 104 ملاحظة. تقوم نماذج تحليل السلاسل الزمنية بالتحقيق في البيانات التي تتميز بديناميات زمنية وتهدف إلى التنبؤ بقيمة المتغير المستهدف لفترة مستقبلية واحدة أو أكثر.
- **قواعد الترابط Association rules** تُستخدم قواعد الارتباط ، المعروفة أيضًا باسم مجموعات التقارب ، لتحديد الارتباطات المثيرة للاهتمام والمتكررة بين مجموعات سجلات مجموعة البيانات. على سبيل المثال ، من الممكن تحديد المنتجات التي يتم شراؤها معًا في معاملة واحدة ومدى تكرار ذلك. تلجأ الشركات في صناعة البيع بالتجزئة إلى قواعد الارتباط لتصميم ترتيب المنتجات على الرفوف أو في الكتالوجات. تُستخدم التجميعات حسب العناصر ذات الصلة أيضًا للترويج للبيع العابر أو لابتكار مجموعات من المنتجات والخدمات والترويج لها.
- **العنقدة Clustering** يشير مصطلح الكتلة إلى مجموعة فرعية متجانسة موجودة داخل مجموعة سكانية. لذلك تهدف تقنيات التجميع إلى تقسيم السكان غير المتجانسين إلى عدد معين من المجموعات الفرعية المكونة من الملاحظات التي تشترك في خصائص متشابهة ؛ الملاحظات المدرجة في مجموعات مختلفة لها سمات مميزة. على عكس التصنيف ، في التجميع لا توجد فئات محددة مسبقًا أو أمثلة مرجعية تشير إلى الفئة المستهدفة ، بحيث يتم تجميع الكائنات معًا بناءً على تجانسها المتبادل. في بعض الأحيان ، يمثل تحديد المجموعات مرحلة أولية في عملية التنقيب عن البيانات ، ضمن تحليل البيانات الاستكشافية. قد يسمح بمعالجة البيانات المتجانسة باستخدام القواعد والتقنيات الأكثر

ملاءمة وتقليل حجم مجموعة البيانات الأصلية ، حيث يمكن تطوير أنشطة التنقيب عن البيانات اللاحقة بشكل مستقل في كل مجموعة محددة.

- **الوصف والتصور Description and visualization:** يتمثل الغرض من عملية التنقيب عن البيانات في بعض الأحيان في توفير تمثيل بسيط وموجز للمعلومات المخزنة في مجموعة بيانات كبيرة. على الرغم من أنه ، على عكس قواعد التجميع والارتباط ، لا يتبع التحليل الوصفي أي تجميع أو تقسيم معين للسجلات في مجموعة البيانات ، إلا أن الوصف الفعال والموجز للمعلومات مفيد للغاية ، لأنه قد يقترح تفسيرات محتملة للأنماط المخفية في البيانات ويؤدي إلى فهم أفضل للظواهر التي تشير إليها البيانات. لاحظ أنه ليس من السهل دائمًا الحصول على تصور مفيد للبيانات. ومع ذلك ، فإن جهد التمثيل له ما يبرره من الإيجاز الملحوظ للمعلومات التي يتم الحصول عليها من خلال مخطط جيد التصميم.

5.5 الملاحظات والقراءات

هناك العديد من الكتب المخصصة لتنقيب البيانات ، مثل Hand et al. 2001 ، هان وكامبر (2005) ، هاستي وآخرون. (2001) ، ويتن وفرانك (2005) وبايل (2003). يركز مجلد Miller and Han (2000) على تحليل البيانات الجغرافية ، بينما يناقش Berry and Linoff (2002) استخراج المعلومات من الويب. تم تقديم دراسة استقصائية في Powell 2001)). وتجدر الإشارة أيضًا إلى الكتب التالية التي تقدم التطبيقات ذات الصلة بالتنقيب في البيانات: ، Berson and Smith (1997) ، (2004 ، 1999) Berry and Linoff Berson et al.

(1999) ، Parr Rud (2000) و Giudici (2003). تم تقديم مسح لطرق التحسين المطبقة في مجال التنقيب عن البيانات في Bradley et al. 1999).

References المصادر

- Adamo J. (2001). Data mining for association rules and sequential patterns: sequential and parallel algorithms. Springer.
- Agrawal R., Shafer J. (1998). Parallel mining of association rules. IEEE Transactions on Knowledge and Data Engineering, 8, 962–969.
- Agrawal R., Aggarwal C., Prasad V. (2001). A tree projection algorithm for generation of frequent itemsets. Journal of Parallel and Distributed Computing, 61, 350–371.
- Aggarwal C., Yu P. (1998). Mining large itemsets for association rules. Data Engineering Bulletin, 21, 23–31.
- Agrawal R., Srikant R. (1994). Fast algorithms for mining association rules in large databases. In: VLDB '94: Proceedings of the 20th International Conference on Very Large Data Bases, 487–499. Morgan Kaufmann.
- Agrawal R., Srikant R. (1995). Mining sequential patterns. In: P. Yu et A. Chen (eds.), ICDE '95: Proceedings of the Eleventh International Conference on Data Engineering, 3–14. IEEE Computer Society.
- Agrawal R., Srikant R. (1997). Mining generalized association rules. Future Generation Computer Systems, 13, 161–180.
- Agrawal R., Imielinski T., Swami A. (1993a). Database mining: A performance perspective. IEEE Transactions on Knowledge and Data Engineering, 5, 914–925.
- Agrawal R., Imielinski T., Swami A. (1993b). Mining association rules between sets of items in large databases. In: SIGMOD '93: Proceedings of the 1993 ACM SIGMOD international conference on Management of data, 207–216. ACM Press.
- Agrawal R., Mannila H., Srikant R., Toivonen H., Verkamo A. (1996). Fast discovery of association rules. In: Advances in knowledge discovery and data mining, 307–328. American Association for Artificial Intelligence.
- Aldenderfer M., Blashfield R. (1985). Cluster Analysis. Sage Publications.
- Allison P. (1998). Multiple Regression: A Primer. Pine Forge Press.
- Anderberg M. (1973). Cluster Analysis for Applications. Academic Press.
- Anderson J. (1995). An Introduction to Neural Networks. MIT Press.
- Anscombe F. (1973). Graphs in statistical analysis. The American Statistician, 195–199.

- Ayres I. (2007). Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart. Bantam.
- Bakan J. (2005). The Corporation: The Pathological Pursuit of Profit and Power. Free Press.
- Baldi P., Brunak S. (2001). Bioinformatics: the machine learning approach. MIT Press.
- Battistini V., Contini A., Del Prato G., Palopoli G., Valentini D., Vercellis C. (1999).
- L'ottimizzazione della catena logistica integrata: il caso Barilla alimentare. Logistica e Management, 102, 77–86.
- Berry M., Linoff G. (1999). Mastering Data Mining. Wiley.
- Berry M., Linoff G. (2002). Mining the Web: Transforming Customer Data into Customer Value. Wiley.
- Berry M., Linoff G. (2004). Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. Wiley.
- Berson A., Smith S. (1997). Data Warehousing, Data Mining, and OLAP . Mcgraw-Hill.
- Berson A., Smith S., Thearling K. (1999). Building Data Mining Applications for CRM . McGraw-Hill.
- Bertsekas D. (2003). Convex Analysis and Optimization. Athena Scientific.
- Bishop C. (1995). Neural Networks for Pattern Recognition. Oxford University Press.
- Bolloju N., Khalifa M., Turban E. (2002). Integrating knowledge management into enterprise environments for the next generation decision support. Decision Support Systems, 33, 163–176.
- Box G., Jenkins G., Reinsel G. (1994). Time Series Analysis: Forecasting & Control . Prentice Hall.
- Bradley P., Fayyad U., Mangasarian O. (1999). Mathematical programming for data mining: formulations and challenges. INFORMS Journal on Computing, 11, 217–238.
- Breiman L., Friedman J., Olshen R., Stone C. (1984). Classification and Regression Trees. Chapman & Hall.
- Breslow L., Aha D. (1997). Simplifying decision trees: A survey. Knowledge Engineering Review , 12, 1–40.
- Brockwell P., Davis R. (2002). Introduction to Time Series and Forecasting. Springer.

- Bruhn M. (2002). Relationship Marketing: Management of Customer Relationships. Pearson.
- Burges C. (1998). A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2, 121–167.
- Cadez I., Heckerman D., Smyth P., Meek C., White S. (2003). Model-based clustering and visualization of navigation patterns on a web site. Data Mining and Knowledge Discovery, 7, 399–424.
- Charnes A., Cooper W., Rhodes E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429–444.
- Chatfield C. (2003). The Analysis of Time Series: An Introduction. Chapman & Hall.
- Cherkassky V., Mulier F. (1998). Learning from data, concepts, theory and methods. Wiley.
- Chopra S., Meindl P. (2003). Supply Chain Management. Prentice Hall.
- Clemen R. (1997). Making Hard Decisions: An Introduction to Decision Analysis. Duxbury Press.
- Cleveland W. (1993). Visualizing Data. Hobart Press.
- Cooper W., Seiford L., Tone K. (2000). Data Envelopment Analysis. A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Springer.
- Crawley M. (2005). Statistics: An Introduction using R. Wiley.
- Cristianini N., Shawe-Taylor J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.
- Cristianini N., Shawe-Taylor J. (2004). Kernel Methods for Pattern Analysis. Cambridge University Press.
- Davenport T., Harris J. (2007). Competing on Analytics: The New Science of Winning. Harvard Business School Press.
- Dhar V., Stein R. (1997). Intelligent Decision Support Methods: The Science of Knowledge. Prentice-Hall.
- Domingos P., Pazzani M. (1997). On the optimality of the simple bayesian classifier under zero-one loss. Machine Learning, 29, 103–130.
- Drucker H., Vapnik V., Wu D. (1999). Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 10, 1048–1054.
- Duda R., Hart P., Stork D. (2001). Pattern Classification. Wiley.

- Dumoulin A., Vercellis C. (2000). Tactical models for hierarchical capacitated lot-sizing problems with setups and changeovers. International Journal of Production Research, 38, 51–67.
- Durbin J., Koopman S. (2001). Time Series Analysis by State Space Methods. Oxford University Press.
- Dyche J. (2000). e-Data: Turning Data Into Information with Data Warehousing. Addison-Wesley.
- Dyche J. (2001). The CRM Handbook: A Business Guide to Customer Relationship Management. Addison-Wesley.
- Egan J. (2001). Relationship Marketing: Exploring Relational Strategies in Marketing. Prentice Hall.
- Eliashberg J., Lilien G. (1993). Marketing. Elsevier Science.
- Elmaghraby S. (1978). Activity networks. Wiley.
- Everitt B., Landau S., Leese M. (2001). Cluster Analysis. Arnold Publishers.
- Evgeniou T., Poggio T., Pontil M., Verri A. (2002). Regularization and statistical learning theory for data analysis. Journal of Computational Statistics and Data Analysis, 38, 421–432.
- Farrell M. (1957). The measurement of productive efficiency. Journal of the Royal Statistical Society, 120, 253–290.
- Fausett L. (1994). Fundamentals of Neural Networks. Prentice Hall.
- Fayyad U., Grinstein G., Wierse A. (eds.) (2002). Information visualization in data mining and knowledge discovery. Morgan Kaufmann.
- Fisher D. (1996). Iterative optimization and simplification of hierarchical clusterings. Journal of Artificial Intelligence Research, 4, 147–179.
- Fisher R. (1936). The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, 179–188.
- Fraley C., Raftery A. (2002). Model-based clustering, discriminant analysis, and density estimation. Journal of the American Statistical Association, 97, 611–631.
- Freitas A. (2000). Understanding the crucial differences between classification and discovery of association rules - a position paper. SIGKDD Explorations Newsletter, 2, 65–69.
- Fukunaga K. (1990). Introduction to Statistical Pattern Recognition. Academic Press.

- Fumero F., Vercellis C. (1997). Integrating distribution, lot-sizing and machine loading via lagrangean relaxation. International Journal of Production Economics, 49, 45–54.
- Fumero F., Vercellis C. (1999). Synchronized development of production, inventory and distribution schedules. Transportation Science, 33, 330–340.
- Giovinazzo W. (2002). Internet-Enabled Business Intelligence. Prentice Hall.
- Giudici P. (2003). Applied Data Mining: Statistical Methods for Business and Industry. Wiley.
- Gorry G., Scott Morton M. (1971). A framework for management information systems. Sloan Management Review , 13, 55–70.
- Graves S., De Kok A. (2003). Supply Chain Management: Design, Coordination and Operation. Elsevier.
- Graves S., Rinnooy Kan A., Zipkin P. (2002). Logistics of Production and Inventory. North Holland.
- Halkidi M., Batistakis Y., Vazirgiannis M. (2002a). Cluster validity methods: part i. SIGMOD Record , 31, 40–45.
- Halkidi M., Batistakis Y., Vazirgiannis M. (2002b). Cluster validity methods: part ii. SIGMOD Record , 31, 19–27.
- Hamilton J. (1994). Time Series Analysis. Princeton University Press.
- Han J., Fu Y. (1995). Discovery of multiple-level association rules from large databases. In: VLDB '95: Proceedings of the 21th International Conference on Very Large Data Bases, 420–431. Morgan Kaufmann.
- Han J., Kamber M. (2005). Data Mining: Concepts and Techniques. Morgan Kaufmann.
- Hand D., Mannila H., Smyth P. (2001). Principles of Data Mining. MIT Press.
- Hartigan J. (1975). Clustering Algorithms. Wiley.
- Harvey A. (1993). Time Series Models. Harvester Wheatsheaf.
- Hastie T., Tibshirani R., Friedman J. (2001). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Haykin S. (1998). Neural Networks: A Comprehensive Foundation. Prentice Hall.
- Hipp J., Guntzer U., Nakhaeizadeh G. (2000). Algorithms for association rule mining a general survey. SIGKDD Explorations Newsletter, 2, 58–64.

- Hogg R., Craig A., McKean J. (2004). Introduction to Mathematical Statistics. Prentice Hall.
- Holt C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Technical report, Carnegie Institute.
- Inmon W. (2002). Building the Data Warehouse. Wiley.
- Jain A., Dubes R. (1988). Algorithms for Clustering Data. Prentice Hall.
- Jain A., Murty M., Flynn P. (1999). Data clustering: A review. ACM Computing Surveys, 31, 264–323.
- Jensen D., Cohen P. (2000). Multiple comparisons in induction algorithms. Machine Learning, 38, 309–338.
- Joachims T. (2002). Learning to Classify Text using Support Vector Machines. Kluwer.
- Jones R. (1980). Maximum likelihood fitting of arma models to time series with missing observations. Technometrics, 20, 389–395.
- Kaufman L., Rousseeuw P. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley.
- Keen P., Scott Morton M. (1978). Decision support systems: an organizational perspective. Addison-Wesley.
- Keys P. (1995). Understanding the process of operational research. Wiley.
- Kimball R. (1996). The Data Warehouse Toolkit. Wiley.
- Kimball R., Ross M. (2002). The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling. Wiley.
- Kimball R., Thornthwaite W., Reeves L., Ross M. (1998). The Data Warehouse Lifecycle Toolkit. Wiley.
- Klein M., Methlie L. B. (1995). Knowledge-based Decision Support Systems with Applications in Business. Wiley.
- Kleinbaum D., Kupper L., Muller K., Nizam A. (1997). Applied Regression Analysis and Multivariable Methods. Duxbury Press.
- Kleinrock L. (1975). Queueing Systems. Wiley-Interscience.
- Kohavi R. (1995). A study on cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence, 1137–1143.

- Kohavi R., John G. (1997). Wrappers for feature subset selection. Artificial Intelligence, 97, 273–324.
- Kudyba S., Hoptroff R. (2001). Data Mining and Business Intelligence: A Guide to Productivity. Idea Group.
- Kulkarni S., Lugosi G., Venkatesh S. (1998). Learning pattern classification - a survey. IEEE Transactions on Information Theory, 44, 2178–2206.
- Lawrence R., Almasi G., Kotlyar V., Viveros M., Duri S. (2001). Personalization of supermarket product recommendations. Data Mining and Knowledge Discovery, 5, 11–32.
- Levine D., Krehbiel T., Berenson M. (2003). Business Statistics - A First Course. Prentice Hall.
- Makridakis S., Wheelwright S. (1989). Forecasting Methods for Management. Wiley Interscience.
- Makridakis S., Wheelwright S., McGee V. (1983). Forecasting: Methods and Applications. Wiley-Hamilton.
- Makridakis S., Wheelwright S., Hyndman R. (1998). Forecasting: Methods and Applications. Wiley.
- Mallach E. (2000). Decision Support and Data Warehouse Systems. McGraw-Hill.
- Malone T., Crowston K., Herman A. (2003). Organizing Business Knowledge: The MIT Process Handbook . The MIT Press.
- Mangasarian O. (1997). Mathematical programming in data mining. Data Mining in Knowledge Discovery, 1, 183–201.
- Mannila H., Toivonen H., Verkamo A. (1994). Efficient algorithms for discovering association rules. In: U. Fayyad et R. Uthurusamy (eds.), AAAI Workshop on Knowledge Discovery in Databases, 181–192. AAAI Press.
- Marshall B., McDonald D., Chen H., Chung W. (2004). Ebizport: collecting and analyzing business intelligence information. Journal of the American Society for information Science and Technology, 55, 873–891.
- Mendenhall W., Sincich T. (2003). A Second Course in Statistics: Regression Analysis. Prentice Hall.
- Mendenhall W., Beaver R., Beaver B. (2000). A Brief Course in Business Statistics. South-Western College Pub.

- Miller H., Han J. (2000). Geographic Data Mining and Knowledge Discovery. Taylor and Francis.
- Mirkin B. (1996). Mathematical Classification and Clustering, volume 11. Kluwer.
- Montgomery D., Peck E., Vining G. (2001). Introduction to Linear Regression Analysis. Wiley-Interscience.
- Moss L., Atre S. (2003). Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley.
- Murtagh F. (1985). Multidimensional Clustering Algorithms. Physica-Verlag.
- Murthy S. (1998). Automatic construction of decision trees from data: a multi-disciplinary survey. Data Mining and Knowledge Discovery, 2, 345–389.
- Murthy S., Kasif S., Salzberg S. (1994). A system for induction of oblique decision trees. Journal of Artificial Intelligence Research, 2, 1–33.
- Nemati H., Steiger D., Iyer L., Herschel R. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. Decision Support Systems, 33, 143–161.
- Nemhauser G., Rinnooy Kan A., Todd M. (1994). Optimization. Elsevier Science.
- Nigam K., McCallum A., Thrun S., Mitchell T. (2000). Text classification from labeled and unlabeled documents using em. Machine Learning, 39, 103–134.
- Orsenigo C., Vercellis C. (2003). Multivariate classification trees based on minimum features discrete support vector machines. IMA Journal of Management Mathematics, 14, 221–234.
- Orsenigo C., Vercellis C. (2004). Discrete support vector decision trees via tabu-search. Journal of Computational Statistics and Data Analysis, 47, 311–322.
- Orsenigo C., Vercellis C. (2006). Rule induction through discrete support vector decision trees. In: E. Triantaphyllou et G. Felici (eds.), Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques, 305–325. Springer.
- Orsenigo C., Vercellis C. (2007). Accurately learning from few examples with a polyhedral classifier. Computational optimization and applications, 38, 235–247.
- Orsenigo C., Vercellis C. (2009). Multicategory classification via discrete support vector machines. Computational Management Science, 1, 101–114.
- Pampel F. (2000). Logistic Regression: A Primer. SAGE Publications.

- Park J., Chen M., Yu P. (1995). An effective hash based algorithm for mining association rules. In: M. Carey et D. Schneider (eds.), Proceedings of the 1995 ACM SIGMOD international conference on Management of data, 175–186. ACM Press.
- Parr Rud O. (2000). Data Mining Cookbook: Modeling Data for Marketing, Risk and Customer Relationship Management. Wiley.
- Pasquier N., Bastide Y., Taouil R., Lakhal L. (1999). Discovering frequent closed itemsets for association rules. In: ICDT '99: Proceeding of the 7th International Conference on Database Theory, 398–416. Springer.
- Pei J., Han J., Mortazavi-Asl B., Zhu H. (2000). Mining access patterns efficiently from web logs. In: PADKK '00: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Current Issues and New Applications, 396–407. Springer.
- Peppers D., Rogers M. (1996). The One to One Future. Currency.
- Peppers D., Rogers M. (2004). Managing Customer Relationships: A Strategic Frame work . Wiley.
- Poirier C., Reiter S. (1999). Supply Chain Optimization. Berrett-Koehler.
- Powell R. (2001). DM Review: A 10 Year Journey. DM Review.
- Pyle D. (1999). Data Preparation for Data Mining. Morgan Kaufmann.
- Pyle D. (2003). Business Modeling and Data Mining. Morgan Kaufmann.
- Quinlan J. (1993). C4.5: Programs for machine learning. Morgan Kaufmann Publishers.
- Ramoni M., Sebastiani P. (2001). Robust bayes classifiers. Artificial Intelligence, 125, 209–226.
- Rasmussen N., Goldy P., Solli P. (2002). Financial Business Intelligence. Trends, Technology, Software Selection, and Implementation. Wiley.
- Raudys S. (2001). Statistical and Neural Classifiers. Springer.
- Ray S. (2004). Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research. Cambridge University Press.
- Ripley B. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.
- Romesburg C. (1984). Cluster Analysis for Researchers. Life Time Learning.
- Rosenblatt F. (1962). Principles of Neurodynamics: Perceptrons and the theory of brain mechanisms. Spartan Books.

- Ryan T. (1997). Modern Regression Methods. Wiley.
- Safavian S., Landgrebe D. (1998). A survey of decision tree classifier methodology. IEEE Transactions on Systems, Man and Cybernetics, 22, 660–674.
- Schafer J., Konstan J., Riedl J. (2001). E-commerce recommendation applications. Journal of Data Mining and Knowledge Discovery, 5, 115–153.
- Schölkopf B., Smola A. (2001). Learning with Kernels, Support Vector Machines, Regularization, Optimization, and Beyond . MIT Press.
- Schurmann J. (1996). Pattern classification, a unified view of statistical and neural approaches. Wiley.
- Scott Long J. (1997). Regression Models for Categorical and Limited Dependent Variables. SAGE Publications.
- Shapiro J. (2000). Modeling the Supply Chain. Duxbury Press.
- Shim J., Warkentin M., Courtney J., Power D., Sharda R., Carlsson C. (2002). Past, present, and future of decision support technology. Decision Support Systems, 33, 111–126.
- Silverstein C., Brin S., Motwan R. (1998). Beyond market baskets: Generalizing association rules to dependence rules. Data Mining and Knowledge Discovery, 2, 39–68.
- Simchi-Levi D., Kaminsky P., Simchi-Levi E. (2002). Designing and Managing the Supply Chain. McGraw-Hill/Irwin.
- Simchi-Levi D., Wu S., Shen Z. (2004a). Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era. Springer.
- Simchi-Levi D., Chen X., Bramel J. (2004b). The Logic of Logistics: Theory, Algorithms, and Applications for Logistics and Supply Chain Management. Springer.
- Simon A., Shaffer S. (2001). Data warehousing and business intelligence for e-commerce. Morgan Kaufmann.
- Simon H. (1969). The sciences of the artificial . The MIT Press.
- Simon H. (1977). The New Science of Management Decision. Prentice Hall.
- Skinner D. (1999). Introduction to Decision Analysis. Probabilistic.
- Snapper J. (1998). Responsibility for computer-based decisions in health care. In: K. Goodman (ed.), Ethics, Computing and Medicine: Informatics and the Transformation of Health Care, 43–56. Cambridge University Press.
- Snedecor G., Cochran W. (1989). Statistical Methods. Iowa State University Press.

- Spaeth H. (1980). Cluster Analysis Algorithms for Data Reduction and Classification of Objects, volume 4. Ellis Horwood.
- Sprague R. (1980). A framework for research on decision support systems. In: G. Fick et R. Sprague (eds.), Decision support systems: issues and challenges, 1–22. Pergamon Press.
- Thanassoulis E. (2001). Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text. Cambridge University Press.
- Toivonen H. (1996). Sampling large databases for association rules. In: T. Vijayaraman, A. Buchmann, C. M., et N. Sarda (eds.), In Proc. 1996 Int. Conf. Very Large Data Bases, 134–145. Morgan Kaufman.
- Tufte E. (1983). The Visual Display of Quantitative Information. Graphics Press.
- Tukey J. (1977). Exploratory Data Analysis. Addison-Wesley.
- Turban E., Aronson J., Liang T. (2005). Decision Support Systems and Intelligent Systems. Prentice Hall.
- Vapnik V. (1996). The Nature of Statistical Learning Theory. Springer.
- Vapnik V. (1998). Statistical learning theory. Wiley.
- Velleman P., Hoaglin D. (1981). The ABC's of EDA: Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press.
- Vercellis C. (2008). Ottimizzazione. Modelli, metodi, applicazioni . McGraw-Hill.
- Weiss N. (2001). Introductory Statistics. Addison Wesley.
- Winkler R., Makridakis S. (1983). The combination of forecasts. Journal of the Royal Statistical Society, A, 145, 150–157.
- Winston W., Venkataramanan M. (2002). Introduction to Mathematical Programming: Applications and Algorithms. Duxbury Press.
- Winters P. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324–342.
- Witten I., Frank E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.
- Zaki M. (1999). Parallel and distributed association mining: A survey. IEEE Concurrency, 7, 14–25.
- Zaki M. (2000). Generating non-redundant association rules. In: KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, 34–43. ACM Press.

- Zaki M., Ogihara M. (1998). Theoretical foundations of association rules. In: Proceedings of 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery.
- Zoltners A., Sinha P., Zoltners G. (2001). The Complete Guide to Accelerating Sales Force Performance: How to Get More Sales from Your Sales Force. American Management Association.