

الجمهورية العربية السورية
المعهد العالي للعلوم التطبيقية والتكنولوجيا

ماجستير نظم المعلومات ودعم القرار

أسامة كنعان

أعدت هذه الأطروحة لنيل

شهادة الماجستير في نظم المعلومات ودعم القرار

بناء نموذج لاكتشاف المعرفة في البيانات يعتمد المنهجية الرشيقية سكروم

Building a Knowledge Discovery in Database (KDD) Model
Based on SCRUM Agile Methodology

(SCRUM-BI)

بإشراف

الدكتور صلاح الدّوه جي

الدكتور كادان الجمعة

إهداء

إلى والديّ العزيزين منارتي في كل درج وأساس كل نجاح.

شكر

أتوجه بجزيل الشكر للدكتور صلاح الدّوه جي – عميد كلية الهندسة المعلوماتية بجامعة دمشق وأستاذ محاضر فيها- لكل ما قدمه من دعم ومساعدة وتوجيهات ونصائح خلال فترة إشرافه على إنجاز هذا العمل.

كما أتوجه بالشكر الخالص للدكتور كادان الجمعة –باحث في المعهد العالي للعلوم التطبيقية والتكنولوجيا ودكتور محاضر فيه- لكل ما قدمه من دعم وأفكار ونصائح وإشراف مستمر ومتواصل حتى إنجاز هذا العمل.

أخيراً لا بدّ من توجيه الشكر للكادر التدريسيّ والإداريّ في المعهد العالي للعلوم التطبيقية والتكنولوجيا بشكل عام، وقسم المعلومات بشكل خاص، لجهودهم المبذولة طيلة فترة البحث.

ملخص العمل

تعد تطبيقات ذكاء الأعمال (BI) من التطبيقات المهمة جداً للعديد من الشركات والمؤسسات. تقدم هذه التطبيقات مجموعة من الأدوات والنظم التي تلعب دوراً رئيسياً في عملية التخطيط الاستراتيجي، وتسمح بجمع البيانات، وتخزينها، والوصول إليها، وتحليلها، للمساعدة في عملية صنع القرار.

لا تزال هذه التطبيقات تواجه العديد من الصعوبات والمشاكل، من أبرزها: الجداول الزمنية غير الواضحة، وعدم تلبية توقعات المستخدمين النهائيين، والنتائج غير المفيدة، ومعدلات فشل عالية تتجاوز نسبة ٥٠٪ حالياً. معظم النماذج الحالية المتبعة في بناء وتطوير هذه التطبيقات ذات طابع تقليدي مقارنةً بالتطور السريع في بيئات العمل والتغييرات الديناميكية على مستوى المتطلبات والأولويات فيها.

اكتسبت مجموعة من التقنيات والعمليات المرنة المعروفة باسم "المنهجيات الرشيقة" أهمية كبيرة في مجال تطوير البرمجيات خلال العقدين الماضيين، وذلك لمواجهة تحديات الحصول على برمجيات عالية الجودة في ظل ظروف السوق وبيئة العمل غير المستقرة. تعد المنهجية سكروم أشهر هذه المنهجيات وأكثرها استخداماً، وتقوم على ركائز ثلاث هي: التكيف، والشفافية، والتحقق. تؤمن المنهجية سكروم إطار عمل متكيف ومتجاوب مع التغييرات الداخلية والخارجية، وتعتمد على فرق صغيرة متعددة تعمل بطريقة مترابطة ومكثفة.

نقدم في هذا العمل نموذج جديد لاكتشاف المعرفة في البيانات "SCRUM-BI" يعتمد المنهجية الرشيقة سكروم، للمساعدة في بناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات. يتميز هذا النموذج بأنه أكثر تكيفاً مع التغييرات في المتطلبات والأولويات من جهة، والتطورات المتسارعة في بيئات العمل من جهة أخرى، كما يحسن ويعزز عملية الحصول على المعرفة ومشاركتها، مما يساهم في دعم عملية اتخاذ القرارات الاستراتيجية.

جرى اختبار وتقييم النموذج باستخدام حالة دراسية على قطاع شركات الاتصالات في سوريا.

ABSTRACT

Business Intelligence applications are vital important for many organizations. These kind of applications provide tools and systems that play a key role in the strategic planning, and allow gathering, store, access, and analyzing data to make decision-making more successful.

These applications are facing many problems and difficulties, such as unclear and undefined schedules, failure to meet user expectations, not all results are useful, and a high failure rates exceeding 50%. Most of current basic models in "BI" applications building and development are traditional in face of evolving business environments, and the dynamic changes on requirements and priorities on it.

Over the past two decades, a number of adaptive techniques and processes, known as "Agile methodologies", have gained much importance in IT and software development. Agile methods address the challenge of delivering high quality software products in unstable market and business conditions. SCRUM is the most famous and commonly used of these methodologies, it based on the three pillars: transparency, adaptation, and inspection. It provides an adaptive and responsive framework with the internal and external changes, and depends on multiple small teams working interdependent and intensive manner.

In this work, we are proposing a new model for knowledge discovery in database (KDD) named "**SCRUM-BI**". It based on SCRUM agile methodology to enhance the way of building Business Intelligence and Data Mining applications. This model characterized as more adaptive to the changing requirements, priorities and rapidly evolving business environments. SCRUM-BI Also improves and enhances the process of knowledge obtaining and sharing, which contributes to support strategic decision-making.

The model was validated using a case study on the telecommunications sector in Syria.

فهرس المحتويات

III.....	إهداء
V	شكر
VII.....	ملخص العمل
IX.....	Abstract
XI.....	فهرس المحتويات
XIX	فهرس الأشكال التوضيحية
XXI	فهرس الجداول
XXIII	مسرد المصطلحات
١	الفصل الأول مقدمة والإطار العام للمسألة
١	١ . الإطار العام
١	١-١ . المعرفة وذكاء الأعمال
٣	٢ . دوافع البحث
٤	٣ . إشكالية البحث
٦	٤ . لمحة عن الحل المقترح
٨	١-٤ . الفرضيات
٨	٢-٤ . المساهمات
١٠	٥ . مخطط البحث
XI	

الفصل الثاني ذكاء الأعمال (BI) والمنهجيات الرشيقة	١١
١. مقدمة	١١
٢. ذكاء الأعمال (BI)	١١
١-٢. تطور مفهوم ذكاء الأعمال	١٢
٢-٢. الهدف من ذكاء الأعمال	١٢
٣-٢. بنية تطبيقات ذكاء الأعمال	١٣
٤-٢. الأدوات الأساسية في تطبيقات ذكاء الأعمال	١٤
١-٤-٢. مستودع البيانات (DW)	١٤
٢-٤-٢. أسواق البيانات (DMs)	١٤
٣-٤-٢. التنقيب في البيانات	١٤
٤-٤-٢. البيانات المترقعة	١٤
٣. المنهجيات الرشيقة (Agile Methodology)	١٤
SCRUM. ١-٣	١٦
Adaptive Software Development (ASD). ٢-٣	١٧
Extreme Programming (XP). ٣-٣	١٧
Crystal method family. ٤-٣	١٨
Dynamic Systems Development Method (DSDM). ٥-٣	١٩
Feature-Driven Development (FDD). ٦-٣	٢٠
Agile Modeling (AM). ٧-٣	٢١
٤. مقارنة بين المنهجيات الرشيقة وملاءمتها لتطبيقات ذكاء الأعمال	٢٢
٥. لماذا المنهجية الرشيقة سكروم	٢٥
٦. خاتمة	٢٦

الفصل الثالث الدّراسة المرجعيّة..... ٢٧

١. مقدمة ٢٧

٢. تصنيف نماذج اكتشاف المعرفة..... ٢٧

١-٢. نماذج اكتشاف المعرفة التقليديّة ٢٨

٢-٢. نماذج اكتشاف المعرفة الهجينة..... ٢٨

٣. المعاملات التي بني عليها إطار المقارنة ٣٠

١-٣. النهج المتبع ٣٠

٢-٣. نوع النموذج والأنظمة التجارية الدّاعمة ٣٠

٣-٣. مخازن البيانات ٣١

٤-٣. طبيعة العمليات ٣١

٥-٣. طبيعة الموارد البشريّة ٣١

٦-٣. طبيعة المسارات ٣١

٧-٣. التكيف مع التغيرات ٣٢

٨-٣. البعد المعرفي ٣٢

٩-٣. البعد الاستراتيجي ٣٢

٤. دراسة الأعمال المختارة..... ٣٢

١-٤. اكتشاف المعرفة في قواعد البيانات "KDD" (Fayyad et al, 1996) ٣٤

٢-٤. "SEMMA" by SAS Institute (1997) ٣٧

٣-٤. Cross-Industry-Standard Process for Data Mining "CRISP-DM" by

CRISP-DM (2008) ٤٠

٤-٤. دورة حياة التنقيب في البيانات العامّة "DMLC" (Hofmann, 2009) ٤٥

٥-٤. اكتشاف المعرفة المقادة بالأنطولوجية "ODKD" (Gottgortory, 2010) ٤٩

Adaptive Software Development-Business Intelligence "ASD-BI" ٦-٤	٥٣
.....(Alnoukari, 2012)	
٥٨	
٥٨	
٦٠	
٦٢	
٦٤	
٦٥	
٦٥	
٦٦	
٦٧	
٦٨	
٧١	
٧١	
٧٣	
٧٤	
٧٥	
٧٦	
٧٦	
٧٧	
٧٩	

٨٠ خبير المنهجية ١-٣-٣
٨٠ محلل الأعمال ٢-٣-٣
٨٠ محلل البيانات ٣-٣-٣
٨٠ مهندس البيانات ٤-٣-٣
٨٠ المنقب في البيانات ٥-٣-٣
٨١ خبير مجال العمل ٦-٣-٣
٨١ مهندس المعرفة ٧-٣-٣
٨١ المدير الاستراتيجي ٨-٣-٣

٤. خصائص النموذج SCRUM-BI ٨٢

٨٢ ١-٤. التكيّف مع التغييرات في بيئة العمل والمتطلبات
٨٣ ٢-٤. البعد المعرفي للنموذج SCRUM-BI
٨٣ ٣-٤. البعد الاستراتيجي للنموذج SCRUM-BI

٥. خاتمة ٨٤

الفصل الخامس الحالة الدراسية ٨٥

١. مقدمة ٨٥

٢. الإطار العام للنظام ٨٦

٨٦ ١-٢. هدف النظام BITS
٨٧ ٢-٢. الوظائف الأساسية للنظام BI-TS
٨٧ ٣-٢. مدخلات النظام BI-TS
٨٧ ٤-٢. مخرجات النظام BI-TS
٨٨ ٥-٢. البيئات البرمجية المستخدمة

٣. مراحل العمل ٨٨

٨٩	١-٣ . فهم بيئة العمل
٩٠	٢-٣ . فهم البيانات
٩٢	٣-٣ . وضع الأهداف وتحديد الفرضيات
٩٤	٤-٣ . اختيار البيانات
٩٤	٥-٣ . النمذجة والتقيب في البيانات
٩٦	٦-٣ . التقييم
٩٧	٧-٣ . النشر
١٠٤	٤ . خاتمة
١٠٥	الفصل السادس الخاتمة والآفاق المستقبلية
١٠٧	المراجع
١١٣	ملحق ١ نماذج اكتشاف المعرفة الأخرى والنماذج الأكثر استخداماً
١١٧	ملحق ٢ المنهجية الرشيقية سكروم SCRUM
١١٧	١ . تطوّر مفهوم سكروم (SCRUM)
١١٨	٢ . بنية المنهجية سكروم
١١٨	١-٢ . الأدوار (Roles)
١١٩	١-٢-١ . مالك المنتج (Product Owner)
١١٩	١-٢-٢ . فريق التطوير (Development Team)
١١٩	١-٢-٣ . مدير سكروم (Scrum Master)
١١٩	٢-٢ . الأحداث (Events)
١٢٠	١-٢-٢ . الشوط (Sprint)
١٢٠	٢-٢-٢ . تخطيط الشوط (Sprint Planning)

١٢١.....	٣-٢-٢ . سكروم اليوميّ (Daily Scrum)
١٢١.....	٤-٢-٢ . استعراض الشوط (Sprint Review)
١٢١.....	٥-٢-٢ . الاسترجاع (Retrospective)
١٢٢.....	٣-٢ . المفاهيم (Artifacts)
١٢٢.....	١-٣-٢ . جعبة المنتج (Product Backlog)
١٢٢.....	٢-٣-٢ . جعبة الشوط (Sprint Backlog)
١٢٢.....	٣-٣-٢ . إضافة المنتج (Product Increment)

فهرس الأشكال التوضيحية

٦	الشكل (١) لمحة عن الحلّ المقترح
١٣	الشكل (٢) بنية تطبيقات ذكاء الأعمال
١٦	الشكل (٣) شوط سكروم "SCRUM Sprint" [9]
١٧	الشكل (٤) مراحل ASD الأساسية [27]
١٨	الشكل (٥) عائلة كريستال من المنهجيات "The Crystal Family of Methodologies" [29]
١٩	الشكل (٦) مراحل عمل المنهجية DSDM [30]
٢٠	الشكل (٧) مراحل عمل دورة حياة تصميم وتطوير مشاريع FDD [35]
٢١	الشكل (٨) خريطة مجموعة الممارسات في المنهجية "النمذجة الرشيقة (AM)" [32]
٢٨	الشكل (٩) نموذج Feldens et al., 1998 [36]
٢٩	الشكل (١٠) Internet_enabled Knowledge Discovery (١٩٩٩) [37]
٣٥	الشكل (١١) نموذج Fayyad et al. ١٩٩٦ [2]
٣٦	الشكل (١٢) نموذج (Feldens et al, 1998) [36]
٣٧	الشكل (١٣) نموذج (Refined KDD Paradigm by Collier et al., 1998) [43]
٣٨	الشكل (١٤) خطوات عمل النموذج (SEMMA, 1997) [40]
٤١	الشكل (١٥) البنية الهرمية لنموذج (CRISP-DM, 2008) [45]
٤١	الشكل (١٦) مراحل دورة حياة النموذج (CRISP-DM, 2008) [41]
	الشكل (١٧) مراحل النموذج دورة حياة التنقيب في البيانات "DMLC" المعتمدة من قبل Hofmann and Tierney, 2007 [47]
٤٦	
٥٠	الشكل (١٨) مراحل ومهام نموذج اكتشاف المعرفة المقادة بالأنطولوجية (ODKD, 2010) [38]
٥٤	الشكل (١٩) بنية النموذج (ASD-BI, 2012) [7]
٦٧	الشكل (٢٠) بنية الأشواط خلال دورة حياة النموذج "SCRUM-BI"
٧٠	الشكل (٢١) بنية النموذج SCRUM-BI
٩١	الشكل (٢٢) مخطط علاقات الكيانات (ERD) لقاعدة البيانات المرحلية
٩٢	الشكل (٢٣) مخطط علاقات الكيانات (ERD) لمستودع البيانات (DW) المستخدم في الحالة الدراسية

- الشكل (٢٤) مخطط علاقات الكيانات (ERD) لمخزن المعلومات والمعرفة (IKR) المستخدم في النظام BI-TS ٩٧
- الشكل (٢٥) بوابة النظام (BI-TS) ٩٨
- الشكل (٢٦) مخطط بياني لعدد الزبائن بالنسبة لمتوسط الفاتورة الشهرية ٩٩
- الشكل (٢٧) مخطط بياني لحجم الإيرادات الشهرية ٩٩
- الشكل (٢٨) مخطط بياني لتصنيف الزبائن بالنسبة للمهنة والجنس ١٠٠
- الشكل (٢٩) مخطط بياني لتصنيف الزبائن بالنسبة لمتوسط الفاتورة الشهرية ١٠٠
- الشكل (٣٠) تحديد عدد الحزم المراد طرحها ١٠١
- الشكل (٣١) أفضل الأسعار لكل حزمة بحسب عدد الزبائن المتوقع اشراكهم في كل منها ١٠١
- الشكل (٣٢) مخطط بياني يظهر تأثيرات الحزم المقترحة ببعضها البعض بالنسبة للزبائن المتوقع اشراكهم في كل منها ١٠٢
- الشكل (٣٣) نموذج توزيع الخدمات ضمن كل حزمة من الحزم المقترحة ١٠٢
- الشكل (٣٤) إيرادات الحزم المقترحة ١٠٣
- الشكل (٣٥) مخطط بياني لعدد الزبائن المتوقع اشراكهم في كل حزمة من الحزم المقترحة ١٠٣

فهرس الجداول

الجدول (١) أهم المنهجيات الرشيقة	١٥
الجدول (٢) مقارنة بين المنهجيات الرشيقة وملاءمتها تطبيقات ذكاء الأعمال (BI)	٢٣
الجدول (٣) نماذج اكتشاف المعرفة السابقة المختارة في دراستنا المرجعية	٣٢
الجدول (٤) سمات النموذج (KDD, 1996) ضمن إطار المقارنة المعرف	٣٥
الجدول (٥) سمات النموذج (SEMMA, 1997) ضمن إطار المقارنة المعرف	٣٩
الجدول (٦) سمات النموذج (CRISP-DM, 2008) ضمن إطار المقارنة المعرف	٤٣
الجدول (٧) سمات النموذج (DMLC, 2009) ضمن إطار المقارنة المعرف	٤٧
الجدول (٨) سمات النموذج (ODKD, 2010) ضمن إطار المقارنة المعرف	٥٢
الجدول (٩) سمات النموذج (ASD-BI, 2012) ضمن إطار المقارنة المعرف	٥٦
الجدول (١٠) ملخص نقاط القوة والضعف في نماذج اكتشاف المعرفة المختارة في دراستنا	٥٩
الجدول (١١) مصفوفة السمات والخصائص المميزة لنماذج اكتشاف المعرفة التي تم اختيارها في دراستنا	٦٠
الجدول (١٢) مصفوفة التقابل بين العمليات والموارد البشرية في النموذج (SCRUM-BI)	٨٢

مسرد المصطلحات

المصطلح باللغة الانكليزية	الاختصار	المصطلح باللغة العربية
Agile Methodology		منهجيات رشيقية
Business Intelligence	BI	ذكاء الأعمال
Customer Relationship Management	CRM	إدارة العلاقات مع الزبائن
Daily Scrum meeting		اجتماع سكروم اليومي
Dashboards		لوحات التحكم
Data Marts	DMs	أسواق البيانات
Data Mining		التنقيب في البيانات
Data Warehouse	DW	مستودع البيانات
Decision Support Systems	DSS	أنظمة دعم القرار
Enterprise Resource Planning	ERP	تخطيط موارد المشاريع
Expert Systems	ES	النظم الخبيرة
Explicit Knowledge		معرفة صريحة
Extract, Transfer, Load	ETL	استخراج البيانات، نقلها، وتحميلها
Generic Data Mining Life Cycle	DMLC	دورة حياة التنقيب في البيانات العامة
Group Decision Support and Mediation Systems	GDSMS	نظم دعم القرارات الجماعية والوساطة
Information and Knowledge Repository	IKR	لمخازن البيانات والمعلومات
Knowledge Discovery in Databases	KDD	اكتشاف المعرفة في قواعد البيانات
Knowledge Discovery Process	KDP	عملية اكتشاف المعرفة
Knowledge Management	KM	إدارة المعرفة
Metadata		البيانات المترفة
Metaphor		استعارة
Ontology Driven Knowledge Discovery	ODKD	اكتشاف المعرفة المقاد بالأنطولوجية
Operational Data Stores	ODS	مخازن البيانات التنفيذية
Product Backlog		جعبة المنتج
Product Increment		إضافة المنتج
Product Owner		مالك المنتج
Scrum Master		مدير سكروم

Scrum Team		فريق سكروم
Sprint		شوط
Sprint Backlog		جعبة الشوط
Sprint Planning		تخطيط الشوط
Sprint retrospective		استرجاع الشوط
Sprint Review		استعراض الشوط
Supply Chain Management	SCM	إدارة سلسلة التوريد
Tacit Knowledge		معرفة ضمنية

الفصل الأول

مقدمة والإطار العام للمسألة

١. الإطار العام

يندرج هذا البحث في إطار اكتشاف المعرفة في البيانات، حيث بدأ مصطلح اكتشاف المعرفة (KD) يندرج في إطار اكتشاف المعرفة من البيانات (Knowledge Discovery in Databases) بالظهور في العام ١٩٨٩ [1]. وعرف فياض عملية اكتشاف المعرفة بأنها "العملية التي تهتم بعملية استخراج المعرفة الضمنية، بما في ذلك كيفية تخزين البيانات والوصول إليها، وكيفية إظهار النتائج وترجمتها، وكيفية دعم العمليات بين الآلة والإنسان نمذجتها" [2]. وهذا يعني ببساطة إن التنقيب في البيانات هو أحد خطوات اكتشاف المعرفة في البيانات (KDD)، "فيعدّ المصطلح التنقيب في البيانات أكثر استخداماً في مجال قواعد البيانات والأعمال، بالمقابل المصطلح اكتشاف المعرفة أكثر شهرةً لدى الباحثين في مجال الذكاء الصناعي. يتم استخدام كلا المصطلحين للتعبير عن عملية البحث عن المعرفة المفيدة ضمن البيانات" [3].

١-١. المعرفة وذكاء الأعمال

تقسم المعرفة لنوعين هما: معرفة ضمنية وأخرى معرفة صريحة، فالمعرفة الضمنية تتسم بأنها شخصية تتعلق بالعقل البشري، ومن الصعب صياغتها أو نمذجتها فتتكون نتيجة الخبرات اليومية وتتأثر بالمعتقدات والرؤى الشخصية. أما المعرفة الصريحة فهي معرفة قابلة للفهم والانتقال وفق لغات مُنهجة ومعتمدة، فيمكن صياغتها باستخدام لغات رسمية بما تتضمنه من جمل قواعديّة وتعابير رياضية وتوصيفات وأدلة وغير ذلك.

تُعرف إدارة المعرفة (KM: Knowledge Management) بأنها "عملية الاستفادة من كامل القيمة المعرفية لرأس المال الفكري، حيث يمكن أن تظهر هذه القيمة من خلال تكوين المعرفة واكتسابها وتطبيقها ومشاركتها" [4].

من هنا تظهر العلاقة بين المعرفة وذكاء الأعمال (BI)، فالهدف من استخدام تقنيات ومناهج ذكاء الأعمال (BI) هو الوصول إلى معارف ومعلومات دقيقة توصلنا بالنهاية لنتائج سليمة وقرارات مناسبة. المعرفة تغني البيانات من خلال جعلها "ذكية" إن أمكن القول، ويمكن أن تكون دليلاً يقود عمليات التنقيب في البيانات، أو في اكتشاف المعرفة في البيانات (KDD). "لا يمكن أن يعمل التنقيب في البيانات دون معرفة" [5].

طرح آخرون عدة أفكار حول العلاقة بين اكتشاف المعرفة وإدارتها من جهة وذكاء الأعمال (BI) من جهة أخرى، حيث وضع McKnight اكتشاف المعرفة وإدارتها تحت إطار تطبيقات ذكاء الأعمال، من خلال مشاركة الذكاء بين الموظفين بطريقة فعالة للقيام بعدة وظائف تساهم في استمرارية عمل المؤسسة، حيث تنظم المعرفة من خلال تطبيقات ذكاء الأعمال (BI) [6]. بالمقابل رأى Haimila أنّ إدارة المعرفة تساعد تطبيقات ذكاء الأعمال، فهو يستشهد باستخدام تطبيقات ذكاء الأعمال (BI) من قبل وكالات تطبيق القانون كوسيلة للاستفادة -بأقصى حد ممكن- من البيانات التي تم جمعها بالشكل الذي يساعد في اتخاذ قرارات أسرع وأفضل، لأنها تُمكن من النفاذ إلى البيانات لمعرفة الاتجاهات والإحصاءات والخصائص المشتركة للجرائم. فيما نوه Cook and Cook أنّ الكثير من الناس ينسى أنّ مفاهيم إدارة المعرفة وذكاء الأعمال (BI) متجذرة على حد سواء في النظريات الخاصة ببناء تطبيقات إدارة الأعمال وممارستها.

مما سبق يمكننا استخلاص النتائج التالية حول العلاقة بين تطبيقات ذكاء الأعمال (BI) ومفهوم اكتشاف المعرفة وإدارتها [7]:

- تركز تطبيقات ذكاء الأعمال (BI) على المعرفة الصريحة، بينما مفهوم اكتشاف المعرفة وإدارتها يجمع بين كل من المعرفة الصريحة والمعرفة الضمنية.
- يعزز كلا المفهومين عمليات الفهم والتعلم واتخاذ القرارات.
- يوضح لنا الفهم العميق للتكامل بين المفهومين أنّ ذكاء الأعمال (BI) هو جزء من اكتشاف المعرفة في البيانات (KDD) وإدارتها.
- يهدف كلا المفهومين إلى تطوير أداء العمل، وذلك من خلال اتخاذ القرارات السليمة في الوقت المناسب، فالغرض الأساسي من تطبيقات ذكاء الأعمال (BI) تحسين الأداء ودعم عمليات صنع القرار.

- تتمّ تطبيقات ذكاء الأعمال (BI) بأنظمة دعم القرار (DSS) وتتعامل مع المنافسين والزبائن، فيما تتمّ إدارة المعرفة بالموظفين، وعملهم، وخبراتهم.

في النهاية يمكننا القول أنّ "المعرفة جزء لا يتجزأ من تطبيقات ذكاء الأعمال (BI) وتدعم عملية صنع القرار" [7].

٢. دوافع البحث

أصبح ذكاء الأعمال (BI) وتطبيقاته واحداً من أهم أطر العمل في مجال صناعة البرمجيات بما تقدمه من أدوات تساعد الشركات في تنظيم ممتلكاتها المعنوية من بيانات ومعلومات، حتى نستطيع القول أنّه أصبح إطار عمل إلزامي وأساسي في مجال الاقتصاد المعرفي، فهو يقدم الدعم لعملية صنع القرار من خلال اكتشاف المعرفة في البيانات. تمر عملية اكتشاف المعرفة في البيانات بالخطوات التالية: الوصول للبيانات، المعالجة، المكاملة، التحويل، الاكتشاف، التعلّم [8]. وقد عرّف Dhar كثافة الذكاء "بكمية المعلومات المفيدة لدعم عملية اتخاذ القرار، التي يحصل عليها صانعو القرار من خلال الأنظمة التحليلية" [8]. يبقى التحدي الأكبر للشركات والمؤسسات في الاستفادة من هذه التطبيقات في اتخاذ القرارات السليمة في الوقت المناسب في ظل الواقع الزاهن من المنافسة، وما تواجهه من صعوبات بسبب: التطور السريع لبيئات العمل والأسواق، والتغيرات الديناميكية المستمرة على مستوى المتطلبات والأولويات، ووجود العديد من أصحاب المصلحة المشاركين في العمل، بالإضافة للجداول الزمنية غير الواضحة، وتعدد الأساليب والنظم لاستخراج البيانات من مصادرها، والمشاكل التي تعترض توافر المعلومات وجودتها. بالمقابل فإنّ معظم المناهج الحالية ذات طابع تقليدي، في ظل نشوء منهجيات جديدة وتغييرها المستمر، وهذه النظريات والمنهجيات التقليدية غير قادرة على مواكبة هذه الثورة وخاصة في مجال الإدارة والأعمال لاعتمادها بشكل رئيسي على الوثائق والتقارير وقوانين صارمة ضمن بيئات العمل متسارعة التغيير. فأصبح من الضروري إيجاد مناهج جديدة قادرة على التعامل مع التطور والتغيرات الديناميكية الحاصلة على مستوى المتطلبات والبيئات المختلفة، والذي نعمل على تحقيقه في هذا البحث، من خلال محاولة تقديم نموذج جديد للمساعدة في بناء تطبيقات ذكاء الأعمال (BI) يتلافى الصعوبات التي تواجهها النماذج الحالية.

٣. إشكالية البحث

تواجه معظم النماذج الحالية - التي تمّ دراسة عدد منها في الفصل الثالث في هذا البحث - فشلاً في التطبيق في ظل واقع الأسواق المتطورة باستمرار، وبيئة العمل المتغيرة على مستوى المتطلبات والأولويات بشكل متطرد [7] [9] [10] [11] [12]. سنحاول فيما يلي تحليل أبرز نقاط الضعف والقصور فيها:

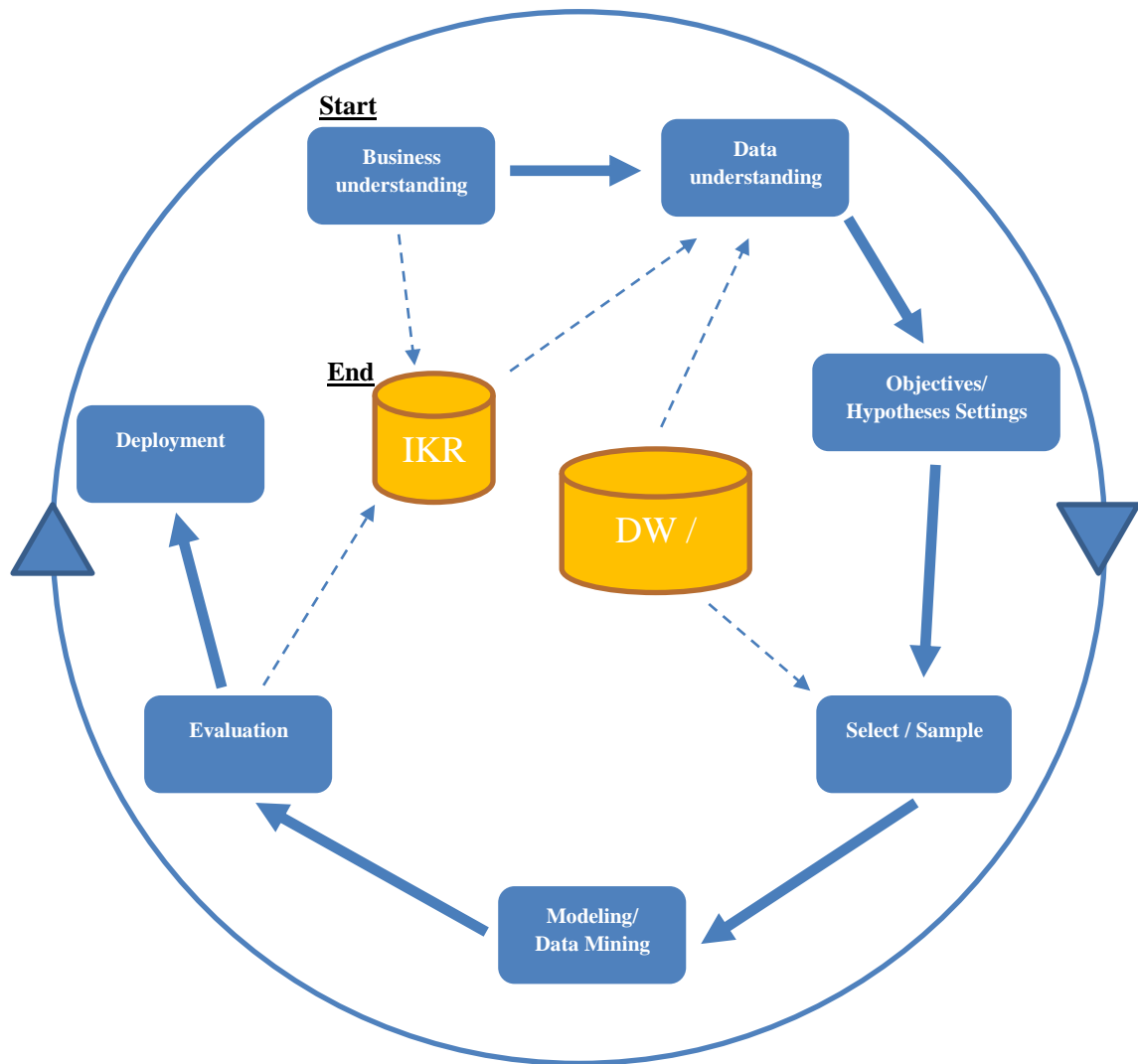
- تتبع معظم النماذج الحالية النهج التقليدي في عملية بناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات: فينتهي المشروع بانتهاء تطوير التطبيق ونشر التقارير النهائية، ولا يخضع المشروع لدورات أخرى بحيث يمكن تصحيح عملية التطوير، والاستفادة من الخبرات المتراكمة في الدورات السابقة من جهة، وردود أفعال المستخدمين النهائيين للتطبيق من جهة أخرى.
- لم تولي بعض النماذج أهمية للبيانات وهيكلتها ومصادرها: فاعتمدت على البيانات المجردة، وأغفلت أهمية الاعتماد على مستودع البيانات وأسواقها (DW/DMS)، والتي تزودنا بشكل أفضل للبيانات، من أجل إجراء عمليات النمذجة والتنقيب في البيانات، وما لذلك من أثر في الحصول على نتائج أفضل.
- لم توضح معظم النماذج آلية تخزين المعرفة المكتشفة، ونقلها، ومشاركتها: للاستفادة منها لاحقاً في الدورات التالية للمشروع، أو في عمليات النمذجة والتنقيب في البيانات لمشاريع أخرى.
- أغفلت معظم النماذج دور الموارد البشرية: فلم تعرّف أدواراً واضحة للأفراد المشاركة في بناء تطبيقات ذكاء الأعمال، والتي تعدّ لاجراً أساسياً في عملية استخلاص المعرفة ومشاركتها بشكل جيّد.
- أغفلت معظم النماذج بعضاً من العمليات الهامة في اكتشاف المعرفة في البيانات (KDD) ومشاركتها: فأغفلت بعض النماذج عملية فهم بيئة العمل، ودور منظور بيئة العمل في توجيه عملية بناء تطبيقات ذكاء الأعمال. كما أغفلت نماذج أخرى عملية وضع الأهداف والفرضيات، والتي تعدّ مرحلة هامة من مراحل بناء تطبيقات ذكاء الأعمال، فتسمح بتحديد أهداف هذه التطبيقات وتوجيهها بالشكل الذي يسهم في تحقيق الخطط الاستراتيجية. ويبقى الأهمّ هو إغفال بعض النماذج عملية النشر، فلم توضح آلية إظهار النتائج وعرضها على المستخدمين النهائيين، للاستفادة منها في تحديد الوضع الحاليّ للمؤسسة أو الشركة، والمساهمة في عملية اتخاذ القرارات، وإيجاد فرص تنافسية جديدة.
- لا تحوي معظم النماذج مسارات داخلية، فلا يمكن مراجعة العمل المنجز حتى الانتهاء من مراحل التطوير ككل في كل دورة للمشروع، وما يتطلبه ذلك من إعادة العمل كاملاً: من إعداد وتحضير للبيانات، وبناء مخازن هذه البيانات، ونقل البيانات لهذه المخازن، وتطبيق آليات النمذجة وخوارزميات التنقيب في البيانات عليها، وبناء التقارير اللازمة لإظهار النتائج وتقييمها.

ولذلك نهدف في بحثنا هذا إيجاد حل للصعوبات والمشاكل التي تواجهها النماذج السابقة، بحيث يتلاقى نقاط الضعف والقصور التي تمّ ذكرها، ويمكننا تلخيص مشكلة البحث بالسؤال التالي:

كيف يمكن بناء نموذج لاكتشاف المعرفة في البيانات (KDD)، يساعد في بناء وتطوير تطبيقات ذكاء الأعمال (BI) بشكل ملائم ومتجاوب مع التغييرات المستمرة والمتسارعة على مستوى المتطلبات وشروط بيئات العمل، ويوجه أهداف هذه المشاريع لتحقيق الخطط الاستراتيجية، ويساهم في دعم عمليات استخلاص المعرفة المكتشفة، وتخزينها، ونقلها، وإظهارها؟

٤. لمحة عن الحل المقترح

نحاول في هذا البحث تقديم حلّ لل صعوبات والمشاكل التي تمّ ذكرها سابقاً، محاولين الاستفادة من نقاط قوة الأبحاث السابقة، وتلافي نقاط الضعف والقصور. تقوم فكرة الحلّ على التكامل والمزاوجة بين نماذج اكتشاف المعرفة تقليدية المنهج والمنهجية الرشيقة لتطوير البرمجيات "المنهجية سكروم (SCRUM)". سنستعرض في هذا الفصل لمحة عامة عن هذا الحلّ موضحين أهم العمليات في دورة حياة هذا الحلّ والمؤشحة بالشكل (١)، والموارد البشرية المضمنة به.



الشكل (١) لمحة عن الحل المقترح

يستخدم الحلّ المقترح في هذا البحث مستودع البيانات وأسواقها (DW\DMs) كمصدر أساسي للبيانات في تطبيقات ذكاء الأعمال (BI) التي يتمّ بناؤها، فيما يستخدم مخزن المعرفة والمعلومات (IKR) لتخزين المعرفة المكتشفة في نهاية كل دورة من دورات الحلّ المقترح وتنظيمها. أمّا الموارد البشرية المتضمنة في هذا الحلّ تتألف بشكل أساسي من: (I) خبير المنهجية. (II) محلل الأعمال. (III) محلل البيانات. (IV) مهندس البيانات. (V) المنقب في البيانات. (VI) خبير مجال العمل. (VII) مهندس المعرفة. (VIII) المدير الاستراتيجي.

تتكون دورة حياة الحلّ المقترح من سبع عمليات متتالية أساسية، وهي عبارة عن قائمة من العمليات التي تمّ اختيارها من عدة نماذج سابقة: (DMLC, 2009)، (CRISP-DM, 2008). وذلك للاستفادة من نقاط قوة كل نموذج وتلافي مواضع القصر فيها. سنستعرض فيما يلي لمحة عن العمليات التي يتكون منها الحلّ المقترح:

١. فهم بيئة العمل (**Business Understanding**): تُمثل نقطة انطلاق الحلّ وواحدة من أهم العمليات فيه، حيث تركز على فهم أهداف ومتطلبات المشروع من وجهة نظر بيئة العمل بالشكل الذي سيساعد في تعريف أهداف وفرضيات المشروع لاحقاً.

٢. فهم البيانات (**Data Understanding**): تُعدّ من العمليات الأساسية في هذا الحلّ إلى جانب عملية فهم بيئة العمل، ويتمّ فيها مجموعة من الأنشطة اللازمة لتحقيق أهداف هذه العملية في فهم البيانات وتوصيفها، على اعتبار أنّ البيانات متوفرة في مستودع البيانات مسبقاً قبل انطلاق المشروع.

٣. تحديد الفرضيات ووضع الأهداف (**Objective/Hypotheses Settings**): تستخدم هذه العملية خرج العمليتين السابقتين لتحديد الفرضيات ووضع أهداف تطبيقات ذكاء الأعمال، بالشكل الذي يسهم في تحقيق الخطط الاستراتيجية للمؤسسات أو الشركات.

٤. اختيار البيانات (**Select/Sample**): تهدف اختيار البيانات اللازمة لتنفيذ خوارزميات التنقيب في البيانات، وجمعها، وصياغتها بالشكل الملائم لهذه الخوارزميات.

٥. النمذجة والتنقيب في البيانات (**Modeling/Data Mining**): تُركز هذه العملية على اختيار تقنيات النمذجة، وخوارزميات التنقيب في البيانات الملائمة للمشاكل وأهداف تطبيق ذكاء الأعمال (BI) التي تمّ تحديدها سابقاً. يعتبر التنقيب في البيانات مكوناً أساسياً لتطبيقات ذكاء الأعمال، حيث يوجد العديد من خوارزميات التنقيب في البيانات: التصنيف (**classification**) والعنقدة (**clustering**) والتجميع (**association**) والتسلسل (**sequencing**) والتنبؤ (**forecasting**).

٦. التقييم (**Evaluation**): تُعتبر واحدة من أهم العمليات في الحلّ المقترح؛ حيث يتمّ التحقق من صحة النموذج المقترح أو خوارزميات التنقيب في البيانات المستخدمة قبل نشر النتائج في العملية اللاحقة، وضمان تحقيق أهداف

تطبيق ذكاء الأعمال (BI) والتي تم وضعها في مرحلة سابقة. فالهدف الرئيسي من هذه المرحلة التحقق من أنّ جميع المسائل الهامة المتعلقة بيئة العمل قد تم أخذها بعين الاعتبار.

٧. النشر (Deployment): تُشكل آخر عملية في دورة حياة الحلّ المقترح، فالنتائج والمعرفة المكتشفة بحاجة لتنظيم، وإدارة، وإيجاد طرق عرض تمكن المستخدم النهائي من فهمها واستخدامها. ليتمّ أخيراً تقييم نتائج الدورة الحالية للمشروع والإعداد للدورة القادمة.

٤-١. الفرضيات

في إطار هذا البحث، يمكننا صياغة الفرضيات التالية:

- ١) تمكن تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات من الوصول لمصادر البيانات المختلفة، وتجميعها، ومكاملتها.
- ٢) تساعد تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات في عمليات تحليل البيانات، ومعالجتها، وتصنيفها بحسب السمات المميزة لها.
- ٣) تقدم تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات للمدراء ومتخذي القرار صورة عن واقع بيئات العمل والأسواق التي يعملون ضمنها، وذلك من خلال المعلومات والمعارف المكتشفة.
- ٤) تساهم تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات في دعم الخطط الاستراتيجية، وذلك من خلال إيجاد فرص تنافسية جديدة في بيئات العمل والأسواق.
- ٥) تشكل تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات أحد المجالات التي يمكن الاستفادة فيها من خصائص ومميزات المنهجيات الرشيقة في التكيف مع التغييرات على مستوى المتطلبات والأولويات ضمن بيئات العمل المتقلبة والمتطورة باستمرار.

٤-٢. المساهمات

نقدم في هذا العمل نموذجاً جديداً لاكتشاف المعرفة في البيانات (KDD) -النموذج "SCRUM-BI"- يأخذ بعين الاعتبار نقاط الضعف والقوة في النماذج السابقة، ويقدم مجموعة من المساهمات أهمها:

- المساعدة في بناء تطبيقات ذكاء الأعمال (BI) أكثر مرونة وتكيفاً مع التغييرات: وذلك بالاستفادة من خصائص المنهجية الرشيقة سكروم التي تركز على الأشخاص والنتائج وتقليل العمليات بمقابل زيادة التعاون والمشاركة. فالمنهجية سكروم مناسبة جداً عندما تكون المتطلبات متقلبة نتيجة لديناميكية بيئات العمل

والتطور المتسارع فيها حيث من الصعب تطبيق المنهجيات التقليدية، "فقد تمّ تطوير المنهجيات الرشيقية

كحل لتطوير البرمجيات في ظل بيئات العمل الغامضة وغير المؤكدة" [13].

- دعم عملية استخلاص المعرفة ومشاركتها: فيركز الحلّ المقترح على الموارد البشرية والتواصل المعرفي فيما بينها من خلال التركيز على أصحاب المصلحة بمشاريع ذكاء الأعمال (BI) والتنقيب في البيانات والاجتماعات المحددة. فالهدف الرئيسي من هذا النموذج هو العمل على زيادة فعالية عملية استخلاص المعرفة، ومشاركتها، وتناقُلها بين مختلف الأشخاص المشاركين في بناء تطبيقات ذكاء الأعمال، من خلال نشاطات اجتماعية كالاتتماعات اليومية وجلسات التقييم.
- مساعدة الشركات والمؤسسات في وضع أهداف تطبيقات ذكاء الأعمال (BI) بشكل يساعد في صياغة خططها الاستراتيجية وتنفيذها: فالنموذج SCRUM-BI يركز بشكل كبير على البعد الاستراتيجي لاستخدام تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، من خلال المدير الاستراتيجي الذي تمّ تضمينه في النموذج بهدف وضع أهداف تطبيقات ذكاء الأعمال (BI) وفرضياتها التي تساعد في صياغة استراتيجية الشركة أو المؤسسة. كما تكشف نتائج تطبيقات ذكاء الأعمال (BI) عن مصادر تنافسية جديدة تساهم في جعل الشركة أكثر مرونة في ظل تغييرات الأسواق.

٥. مخطط البحث

نظمت الأطروحة على عدة فصول، سنقوم فيما يلي باستعراض لمحة عن كل منها:

- الفصل الثاني (ذكاء الأعمال (BI) والمنهجيات الرشيقة): نوضح في هذا الفصل مفهوم ذكاء الأعمال، ونعرّف بأهم المفاهيم والأدوات المرتبطة به، ثم نقوم بعرض لمحة عن أهم المنهجيات الرشيقة، ومميزات كل منها، ونقارن بينها بناءً لخصائصها، ومميزاتها، وملاءمتها لتطبيقات ذكاء الأعمال، لتوضيح سبب اختيار المنهجية سكروم بالمقارنة مع المنهجيات الرشيقة الأخرى.
- الفصل الثالث (الدراسة مرجعية): نقوم في هذا الفصل بتحليل أبرز النماذج المعروفة في هذا المجال، من خلال التعريف باكتشاف المعرفة في البيانات (KDD)، وتصنيف نماذج اكتشاف المعرفة الموجودة، ومن ثمّ نقدم إطاراً للمقارنة بين الأبرز من هذه النماذج لتحليل مواضع الضعف والقوة فيها، لنلخص النتائج في النهاية بمصفوفة معاملات المقارنة بين هذه الأعمال، ونوضح موضع الإسهامات المرجوة في نموذجنا (SCRUM-BI) المقدم في هذا البحث.
- الفصل الرابع (النموذج (SCRUM-BI): نقدم في هذا الفصل نموذجنا المقترح في هذه الدراسة الذي قمنا بتسميته (SCRUM-BI)، موضحين كيف سيتم ملاءمة المنهجية الرشيقة سكروم (SCRUM) في حلّنا المقترح. نفصل بعدها بنية النموذج (SCRUM-BI)، ونوضح أهم الجوانب المتعلقة به وهي: العمليات، والموارد البشرية، ومخازن البيانات. ومن ثمّ نستعرض خصائص هذا النموذج ودوره في التكيف والتجاوب مع المتغيرات الخارجية في المتطلبات وبيئات العمل، وتنظيم الخبرات والمعارف المتراكمة وتناقلها بين أفراد فريق العمل، وصياغة الخطط الاستراتيجية وتحقيقها.
- الفصل الخامس (الحالة الدراسية). نقدم خلال هذا الفصل لمحة عن نظام ذكاء الأعمال (BI) الذي قمنا ببنائه (النظام (BI-TS)، محاولين توضيح التطبيق العملي للنموذج (SCRUM-BI) المقدم في هذا العمل. مبرزين كيف يمكن لتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات مساعدة شركات الاتصالات في تقييم الوضع الحالي للشركة، ودعم القرارات الاستراتيجية في المجالات التسويقية والترويجية لها.
- الفصل السابع (الخاتمة والآفاق المستقبلية).

الفصل الثاني

ذكاء الأعمال (BI) والمنهجيات الرشيقة

١. مقدمة

نعرف في هذا الفصل بأهم المفاهيم المرتبطة بمجال البحث الذي نعمل ضمنه. فنقدم في البداية مفهوم ذكاء الأعمال، وتطور هذا المفهوم، والهدف من تطبيقات ذكاء الأعمال، بالإضافة للتعريف ببنية هذه التطبيقات والأدوات الأساسية ضمنها. ثم نقوم بعرض لمحة عن المنهجيات الرشيقة، وأهم مميزات كل منها، ونقارن بينها بناءً لخصائصها، ومميزاتها، وملاءمتها لتطبيقات ذكاء الأعمال، لتوضيح سبب اختيار المنهجية سكروم بالمقارنة مع منهجيات رشيقة أخرى، وأهمية اعتماد هذه المنهجية في بناء تطبيقات ذكاء الأعمال (BI) ناجحة.

٢. ذكاء الأعمال (BI)

أصبحت تطبيقات ذكاء الأعمال (BI) من التطبيقات الهامة في مجال تكنولوجيا المعلومات والتي تساعد الشركات والمؤسسات في عملية الوصول لممتلكاتها المعنوية من المعلومات والمعرفة الخاصة بها وبعملها وتنظيمها وتطويرها، مما يجعل منها أحد أطر العمل الهامة في قطاع الاقتصاد المعرفي [7].

نقدم فيما يلي تلخيص لأهم المفاهيم والأدوات المستخدمة في بناء تطبيقات ذكاء الأعمال: البنية المعمارية، الأهداف، الأدوات الأساسية (مستودع البيانات، أسواق البيانات، التنقيب في البيانات، ...).

٢-١. تطور مفهوم ذكاء الأعمال

عرّف Luhn الذكاء في إحدى مقالاته العلمية التي نشرها في مجلة IBM: "القدرة على إدراك العلاقات المتبادلة بين الوقائع والحقائق الظاهرة، لتوجيه العمل نحو الهدف المنشود" [14].

أنتج هذا المفهوم خلال العقود اللاحقة العديد من النظم والتطبيقات التي تسعى لتحسين قدرة الشركات والمؤسسات في عملية صنع القرار، ليكون ذكاء الأعمال (BI) وتطبيقاته واحداً من أهم أشكال نظم دعم القرار المتعاقبة في السنوات الأخيرة. تهدف أنظمة دعم القرار إلى مساعدة المدراء لاتخاذ القرارات السليمة في الوقت المناسب [15]، عبر تقديم مجموعة واسعة من تقنيات تحليل القرار، ومحاكاته، ونمذجته، بما في ذلك أشجار القرار وشبكات الاعتقاد. كما أنّها تنطوي على العديد من النظم والأدوات من أهمها: أنظمة دعم القرار (DSS)، ونظم دعم القرارات الجماعية والوساطة (GDSMS)، والنظم الخبيرة (ES)، وذكاء الأعمال (BI) [16].

ومع بداية تسعينيات القرن الماضي ظهر مفهوم ذكاء مستودعات البيانات، ويمكن تعريفه بأنه "مجموعة من العمليات المصممة لاستخراج، وتنظيف، وتمييز البيانات بالشكل الذي يعطي المستخدم فكرة أوضح حول البيانات التي يتعامل معها، وعلاقتها بالموضوع المعالج" [7]. اعتمدت أنظمة دعم القرار فيما بعد بشكل واضح على مستودعات البيانات، وتقنيات OLAP بما تزوّده من أدوات تساعد المستخدمين في تصفح البيانات وتلخيصها بطريقة ديناميكية وفعّالة [17].

في الحقيقة إنّ Howard Dresner عضو مجموعة غارتنير (Gartner group) كان من أوائل من قدموا مصطلح ذكاء الأعمال (BI) كمظلة تضم مجموعة من المفاهيم والطرائق لتحسين عملية صنع القرار باستخدام أنظمة الدعم المرتكزة على الحقائق والوقائع. ليتم إعادة تعريف المصطلح بأنه "مظلة تجمع مجموعة من البنى والأدوات والتطبيقات والخبرات وقواعد البيانات والمنهجيات" [5]. أما Stevan Dedijer فقد اعتبر أنّ إدارة المعرفة نشأت كجزء من التفكير بأساليب ومناهج ذكيّة بالنسبة لبيئة العمل، "فالذكاء أكثر وصفاً من المعرفة، المعرفة ثابتة فيما الذكاء مرّن" [18]. ليظهر فيما بعد العديد من المفاهيم في هذا مجالات عديدة أخرى.

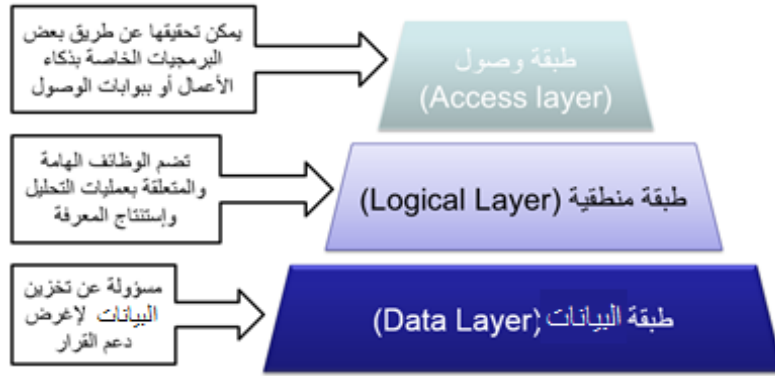
٢-٢. الهدف من ذكاء الأعمال

تهدف حلول ذكاء الأعمال (BI) وتطبيقاتها الوصول للبيانات بمصادرها المختلفة، وتحويلها لمعلومات، ومن ثمّ معرفة، بالشكل الذي يساهم في تحسين عملية صنع واتخاذ القرار. وذلك باستخدام المعرفة المكتشفة من خلال عملية

التنقيب في البيانات، بغرض دعم صانعي القرار من خلال توضيح الواقع الزاهن، أو المساعدة في توقع النتائج المستقبلية [19].

إنّ قدرات الذكاء هي الجزء الأساسي والمعقد وراء أنظمة وتطبيقات ذكاء الأعمال، فعلى النظام أن يحلل نتائج عمليات التنقيب في البيانات في مرحلة لاحقة لمرحلة التنقيب في البيانات، وإظهار النتائج باستخدام منصة إظهار مناسبة. فتقاس حلول ذكاء الأعمال (BI) من خلال القدرة على اشتقاق المعرفة من البيانات، فالتحدي هنا هو القدرة على تعريف النماذج، التوجهات، القواعد، والعلاقات من المعلومات المكتشفة التي يصعب تحليلها البشري بسبب حجمها الهائل [7].

٢-٣. بنية تطبيقات ذكاء الأعمال



الشكل (٢) بنية تطبيقات ذكاء الأعمال

يمكن تجزئة البنية الأساسية لأي نظام من أنظمة ذكاء الأعمال (BI) إلى الطبقات الثلاثة التالية [20] [21]

[17] [7]، والمؤشحة بالشكل (٢):

١. **طبقة البيانات:** تُمكن تخزين البيانات المستخدمة لأغراض دعم القرار بشكلها المهيكلة أو غير المهيكلة، تُخزن البيانات المهيكلة - في العادة - ضمن مخازن البيانات التشغيلية (ODS) أو مستودع البيانات وأسواقها (DW/DMS)، فيما يتم إدارة البيانات غير المهيكلة من خلال أنظمة إدارة المستندات والمحتوى [7]. بجميع الأحوال يتم استخراج البيانات من مصادرها التشغيلية كأنظمة: إدارة العلاقات مع الزبائن (CRM)، تخطيط موارد المشاريع (ERP)، إدارة سلسلة التوريد (SCM)، أو من مصادر خارجية أخرى لتحويلها إلى مستودع البيانات (DW) باستخدام أدوات (ETL).

٢. **طبقة التحليل:** تُوفّر هذه الطبقة القدرة على تحليل البيانات واستخراج المعرفة، بما تتضمنه من عمليات: التنقيب في البيانات، OLAP، التجميع، وغيرها.

٣. طبقة الإظهار: تؤمن مجموعة من البوابات وتطبيقات ذكاء الأعمال.

٢-٤. الأدوات الأساسية في تطبيقات ذكاء الأعمال

يجمع مفهوم ذكاء الأعمال (BI) بين مجموعة من البنى، والأدوات، والتطبيقات، والمنهجيات، بهدف الاستفادة من قواعد البيانات، وتحويلها لمعلومات مفيدة من الممكن أن تساعد متخذي القرار على كافة المستويات، وتزودهم بنظرة أعمق لبيئة العمل التي يعملون ضمنها، نعرف فيما يلي بأهم هذه الأدوات:

٢-٤-١. مستودع البيانات (DW)

هو مخزن فيزيائي يتم مراكمة البيانات فيه من مصادرها المختلفة بشكل مُهيكل ومُنظّم، يزود الشركات بمصدر للبيانات النظيفة، والمهيكلّة، والمصاغة بطريقة معيارية [22].

٢-٤-٢. أسواق البيانات (DMs)

هي مجموعات جزئية من مستودع البيانات (DW) التي تم اختيارها بحسب الحاجة ومتطلبات العمل، وتُشكل طبقة الوصول لمستودع البيانات (DW) إن أمكن القول، والتي تمكن المستخدمين من الوصول للبيانات التي يحتاجونها [23].

٢-٤-٣. التنقيب في البيانات

هي عملية اكتشاف المعارف المفيدة في البيانات، وتلخيص أهم المعلومات التي تزودنا بها هذه البيانات، والتي من الممكن استخدامها في عمليات التوقع والتنبؤ القادمة [24].

٢-٤-٤. البيانات المترفّعة

هي بيانات تصف بيانات أخرى؛ فهي تلخص المعلومات الأساسية حول بيانات أخرى، مما يسهّل عمليات الوصول والمعالجة لهذه البيانات.

٣. المنهجيات الرشيقة (Agile Methodology)

"أسرع، أفضل، أرخص" هذا هو الوعد الذي قدمته المنهجيات الرشيقة، فقد اكتسبت مجموعة من العمليات والتقنيات التكيفية (المعروفة باسم المنهجيات الرشيقة) أهمية كبيرة في مجال تطوير البرمجيات وتكنولوجيا المعلومات خلال

العقدين الماضيين، لمواجهة التحديات المتمثلة في تقديم منتجات برمجية عالية الجودة في ظروف السوق وبيئات العمل غير المستقرة والمتطورة بوتيرة متسارعة.

تتميز المنهجيات الرشيقة بشكل عام بالخصائص ذاتها من حيث التركيز على الأشخاص والنتائج، وتقليل العمليات، وزيادة التعاون، فهي تتبع الفلسفة ذاتها التي تقوم على التركيز على مجموعة من المبادئ [25]. من أهم هذه المبادئ التركيز على:

- الأفراد والتفاعل بينهم عوضاً عن الأدوات والعمليات.
- البرمجيات الفعالة عوضاً عن التوثيق الشامل.
- التعاون مع الزبون عوضاً عن العقود والتفاوض.
- التجاوب مع التغييرات عوضاً عن إتباع خطة ثابتة.

ظهرت العديد من المنهجيات تحت اسم المنهجيات الرشيقة خلال العقدين الماضيين، تعدّ المنهجيات المؤسسية بالجدول (١) التالي من أهم هذه المنهجيات:

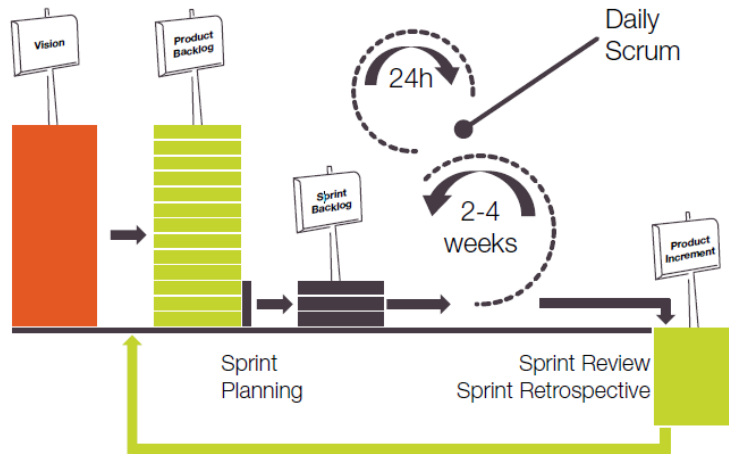
الجدول (١) أهم المنهجيات الرشيقة

اسم المنهجية	الاختصار	عام النشر	الباحثين
SCRUM [26]	SCRUM	2001	Ken Schwaber and Mike Beedle
Adaptive Software Development [27]	ASD	2000	James A. Highsmith
Extreme Programming [28]	XP	1999	Kent Beck
Crystal method family [29]		2002	Alistair Cockburn
Dynamic Systems Development Method [30]	DSDM	1997	Jennifer Stapleton
Feature-Driven Development [31]	FDD	2002	Stephen R. Palmer and John M. Felsing
Agile Modeling [32]	AM	2002	Scott Ambler

تتشارك هذه المنهجيات مجموعة من المبادئ والقواعد كما أسلفنا سابقاً، ولكنها تختلف فيما بينها من حيث العناصر، والهندسية، ومراحل تطوير المنتج، وتوصيف الموارد البشرية، سنقوم فيما يلي بعرض لمحة عن كل من هذه المنهجيات:

٣-١-٣ SCRUM

المنهجية سكروم (SCRUM) هي إحدى أشهر المنهجيات الرشيقة وأكثرها انتشاراً واعتماداً من قبل المؤسسات والشركات حول العالم. "تعمل المنهجية سكروم على تطبيق نظرية التحكم بالعمليات الصناعية في مجال بناء النظم وتطوير البرمجيات بما تقوم عليه من أفكار التكيف، والمرونة، والإنتاجية" [26]. تعتمد المنهجية سكروم (SCRUM) على فرق صغيرة متعددة تعمل بطريقة مكثفة ومتراصة، وتكثرت على أسلوب عمل هذه الفرق بحيث تتطور الأنظمة والتطبيقات في ظل بيئات عمل متغيرة ومتقلبة باستمرار. "تقوم المنهجية سكروم على ركائز ثلاث هي: الشفافية، والتكيف، والتحقق" [33]. يتم تجرئة عمليات تطوير البرمجيات وبنائها في هذه المنهجية ضمن دورات تمتد بين ٧ وحتى ٣٠ يوم، وتدعى هذه الدورات أشواط (Sprints)، كما يظهر الشكل (٣)، كما تنطوي على مجموعة من الأنشطة والاجتماعات اليومية من أجل تحديد أي قصور أو عقبات خلال عمليات التطوير، والعمل على تصحيحها وتلافيها، فتعزف مجموعة من الأحداث هي: (I) الشوط (Sprint). (II) تخطيط الشوط (Sprint Planning). (III) سكروم اليومي (Daily Scrum). (IV) استعراض الشوط (Sprint Review). (V) الاسترجاع (Retrospective). تحديد مسؤوليات فريق العمل - لا يتجاوز ١٠ أفراد غالباً- خلال عمليات تطوير البرمجيات وبنائها من خلال مجموعة من الأدوار هي: (I) مالك المنتج (Product Owner). (II) فريق التطوير (Development Team). (III) مدير سكروم (Scrum Master). يتشارك فريق العمل مجموعة من المفاهيم للحفاظ على الشفافية خلال العمل، وهذه المفاهيم هي: (I) جعبة المنتج (Product Backlog). (II) جعبة الشوط (Sprint Backlog). (III) إضافة المنتج (Product Increment).



الشكل (٣) شوط سكروم "SCRUM Sprint" [9]

٢-٣ Adaptive Software Development (ASD)

المنهجية ASD هي منهجية لتطوير البرمجيات تقوم معظم مبادئها على أبحاث Highsmith السابقة حول أساليب تطوير البرمجيات التكرارية. تنطوي هذه المنهجية على مجموعة من الأساليب لتطوير البرمجيات بشكل متزايد وتكراري مع النشر المستمر للنماذج الأولية، وتبرز - كغيرها من المنهجيات الرشيقية - أهمية تعاون فرق العمل، وعملها بشكل جماعي، لبناء ثقافة تنظيمية قائمة على التكيف. تركز هذه المنهجية على البرمجيات التي يتم تطويرها أكثر من المهام اللازمة لتطويرها. تتضمن هذه المنهجية ثلاث مراحل أساسية تتم خلال دورات متعاقبة كما هو موضح بالشكل (٤)، وهذه المراحل هي: (I) التخمين. (II) التعاون. (III) التعلم.



الشكل (٤) مراحل ASD الأساسية [27]

"تقدم المنهجية ASD عدد قليل جداً من الأنشطة والممارسات اليومية خلال عمليات تطوير البرمجيات، ولا توضح هيكلية فرق العمل، كما لا تفرض وجود أعضاء فريق العمل كالمناهج الرشيقية الأخرى" [34].

٣-٣ Extreme Programming (XP)

المنهجية XP هي إحدى المنهجيات الرشيقية التي تهدف إلى تحسين نوعية البرمجيات والاستجابة لمتطلبات الزبائن المتغيرة. تمكن هذه المنهجية - كغيرها من المنهجيات الرشيقية - من تسليم إصدارات خلال دورات عمل قصيرة، بالشكل الذي يحسن الإنتاجية، ويمكن من وجود نقاط فحص تمكن من التجاوب مع متطلبات الزبائن الجديدة. "تعتبر المنهجية XP فرق العمل، والمدراء، والزبائن، والمطورين جميعاً فريقاً واحداً، وذلك بهدف تقديم برمجيات عالية الجودة" [34]. عرّفت هذه المنهجية أربع قيم أساسية بدايةً في العام ١٩٩٩، هي: (I) التواصل. (II) البساطة. (III) ردود الأفعال. (IV) الشجاعة. ثم تم إضافة القيمة الاحترام في الطبعة الثانية في العام ٢٠٠٠. بناءً على هذه القيم حددت خمس مبادئ أساسية كدليل لعمليات تطوير البرمجيات، هذه المبادئ هي: (I) ردود الأفعال السريعة. (II) افتراض البساطة. (III) التغيير المتزايد. (IV) قبول التغيير. (V) العمل بجودة. تتضمن عمليات تطوير البرمجيات في هذه المنهجية أربع أنشطة أساسية هي: (I) كتابة الرموز المصدري. (II) الاختبار. (III) الاستماع. (IV) التصميم. واستناداً للقيم، والمبادئ، والأنشطة

هذه تم تعريف مجموعة من الممارسات، من أهمها: (I) ألعاب التخطيط (Planning Game). (II) الإصدارات الصغيرة (Small releases). (III) الاستعارة (Metaphor). (IV) بسيطة التصميم (Simple design). (V) إعادة بناء التعليمات البرمجية (Refactoring). (VI) البرمجة الزوجية (Pair programming). (VII) الملكية الجماعية (Collective ownership). (VIII) التكامل المستمر (Continuous integration). (IX) ٤٠ ساعة في الأسبوع (40-hour week). (X) الزبون في الموقع (On-site customer). (XI) كتابة الرماز بناءً للمعايير (Coding standards). تمر دورة حياة المشروع في هذه المنهجية بالمرحل الستة التالية: (I) الاستكشاف. (II) التخطيط. (III) التكرار حتى الإصدار الأول. (IV) الإنتاج. (V) الصيانة. (VI) الموت. تتألف فرق العمل في المنهجية XP من ٢ وحتى ١٢ عضواً، وتعرّف الأدوار الرئيسية خلال مراحل تطوير المشروع بالأدوار التالية: (I) المبرمج. (II) الزبون. (III) مُختبر المشروع. (IV) المدرب. (V) المستشار. (VI) الرئيس الكبير.

٣-٤ . Crystal method family

"عائلة منهجيات كريستال هي مجموعة من المنهجيات المختلفة، ومبادئ التفصيل الخاصة بها لتناسب الطبيعة المتفاوتة للمشاريع المختلفة" [29]. تستخدم هذه العائلة من المنهجيات مجموعة من الألوان لتحديد حجم العمل، وهذه الألوان هي: (I) لا لون. (II) أصفر. (III) برتقالي. (IV) أحمر (v) كستنائي. كما تستخدم مجموعة من الأحرف لتحديد درجة أهمية المشروع، وهي: (I) C: Comfort. (II) D: Discretionary Money. (III) E: Essential Money. (IV) L: Life. وبناءً لهذه الألوان والأحرف تتشكل المصفوفة الموضحة بالشكل (٥)، والتي تمكن من تحديد أي منهجية يجب استخدامها بحسب أهمية المشروع وحجم العمل الذي يتضمنه.

	Clear	Yellow	Orange	Red	Maroon
Life (L)	L6	L20	L40	L80	L200
Essential Money (E)	E6	E20	E40	E80	E200
Discretionary Money (D)	D6	D20	D40	D80	D200
Comfort (C)	C6	C20	C40	C80	C200
	1-6	7-20	21-40	41-80	81-200

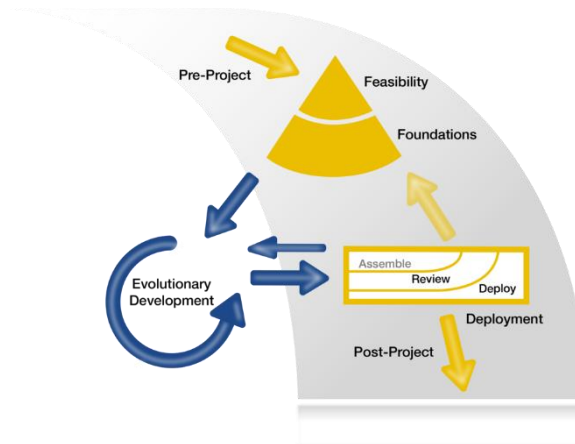
الشكل (٥) عائلة كريستال من المنهجيات "The Crystal Family of Methodologies" [29]

تشارك جميع منهجيات عائلة كريستال بسبع خصائص أساسية، فقد وجد "Cockburn" أنّ هذه الخصائص تشارك فيها المشاريع الناجحة على الأرجح. هذه الخصائص السبعة هي: (I) التسليم المتكرر. (II) التحسينات المنعكسة. (III) التواصل القريب. (IV) السلامة الشخصية. (V) التركيز. (VI) الوصول السهل للمستخدمين الخبراء. (VII) بيئة تقنية تؤمن الاختبارات الآلية، وإدارة التهيئة، والتكامل التكراري. تمر عمليات تطوير وبناء البرمجيات بدورات مستمرة مدة كل منها تتراوح بين شهر وثلاثة أشهر، وينصب التركيز خلالها على التعاون والتفاعل بين الأفراد المشاركين بالعمل.

"لا تقيد منهجيات كريستال استخدام أية ممارسات أو أساليب أخرى، فيمكنها بسهولة اعتماد ممارسات وأساليب المنهجية XP" [34].

٣-٥. Dynamic Systems Development Method (DSDM)

"المنهجية DSDM هي طريقة تم تطويرها بحيث تشكل إطار عمل لتطوير البرمجيات السريع (RAD)" [35]. وضعت النسخة الأولى من هذه المنهجية في المملكة المتحدة في العام ١٩٩٤. الفكرة الأساسية في هذه المنهجية هي تحديد الوقت والموارد، ومن ثم تحديد الوظائف على أساسها، بدلاً من تحدد الوظائف، ومن ثم ضبط الوقت والموارد لتحقيق هذه الوظائف. ظهرت العديد من الإصدارات منذ نشأة هذه المنهجية، وتغيرت علامتها التجارية أكثر من مرة، ويعد الإصدار السادس للعام ٢٠١٤ أحدث هذه الإصدارات. تقوم المنهجية DSDM على ثمانية مبادئ أساسية هي: (I) التركيز على حاجات العمل. (II) التسليم في الوقت المحدد. (III) التعاون. (IV) عدم إهمال الجودة. (V) البناء المتزايد من الأساسيات. (VI) التطوير التكراري. (VII) التواصل المستمر والواضح. (VIII) ثبات السيطرة. تمر دورة حياة تطوير البرمجيات في المنهجية DSDM بالمرحلة الستة - كما يوضح الشكل (٦) - التالية: (I) التحضير للمشروع. (II) دراسة الجدوى والإمكانية. (III) تأسيس. (IV) التطوير التدريجي. (V) النشر. (VI) ما بعد المشروع.



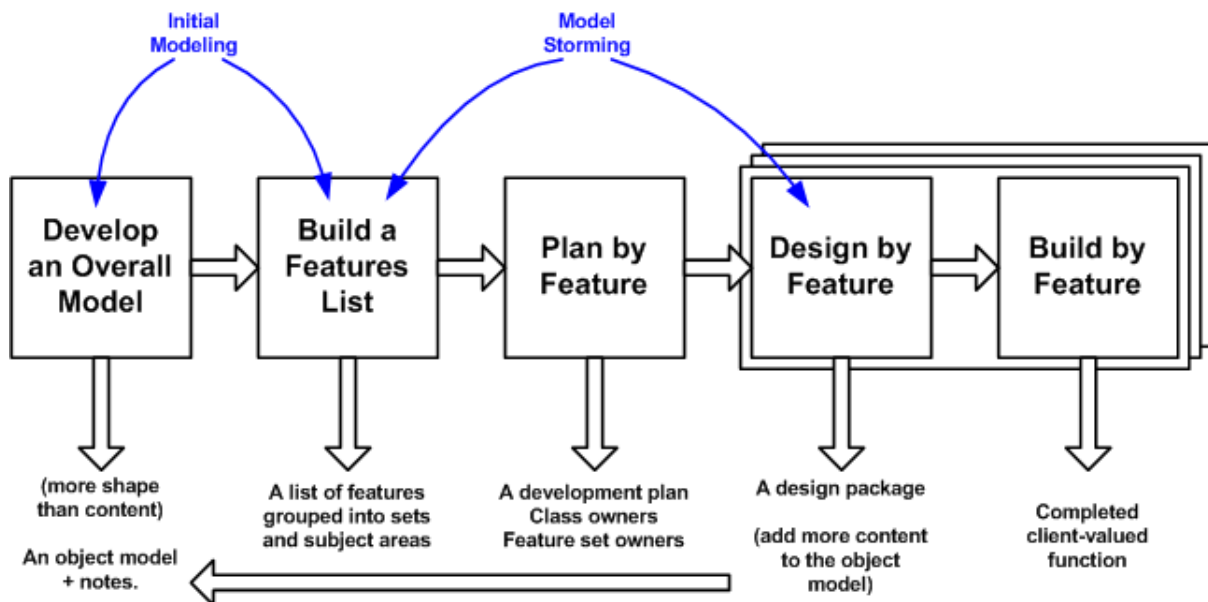
الشكل (٦) مراحل عمل المنهجية DSDM [30]

تعرف هذه المنهجية حوالي ١٣ دوراً للأفراد والمستخدمين، بينما تتكوّن فرق العمل من ٢ إلى ٦ أعضاء، ويمكن وجود أكثر من فريق واحد ضمن المشروع.

على الرغم من وجود أوراق بيضاء (White Papers) للدمج بين DSDM، وUML، وRUP، ولكن الدعم متوفر فقط للشركاء والعملاء مقابل تكلفة سنوية.

٦-٣ . Feature-Driven Development (FDD)

"المنهجية FDD هي طريقة لتطوير البرمجيات موجهة نحو العمليات، بهدف تطوير نظم العمل الحيوية" [35]. تترجم هذه المنهجية بين عدد من أفضل الممارسات الموجودة في القطاع الصناعي، لتشكل نهجاً واضحاً ومتكاملاً. يجعل هذا المزيج من المقومات المنهجية FDD فريدة من نوعها في كل حالة. تشدد هذه المنهجية على مجموعة من الجوانب خلال مراحل العمل، من أهمها: (I) الجودة. (II) التسليم المتكرر. (III) مراقبة تقدم المشروع. تتألف دورة حياة مشاريع FDD من خمس مراحل متتابعة يتم تصميم وتطوير النظم والبرمجيات خلالها، وهذه المراحل هي: (I) تطوير النموذج العام. (II) بناء قائمة الميزات. (III) التخطيط على أساس هذه الميزات. (IV) التصميم على أساس هذه الميزات. (V) البناء على أساس هذه الميزات. كما يوضح الشكل (٧) التالي أنّ المراحل الثلاث الأولى يتم تطويرها بشكل متتابع، فيما تشكل المرحلتين اللاحقتين الجزء التكراري من دورة حياة مشاريع FDD.



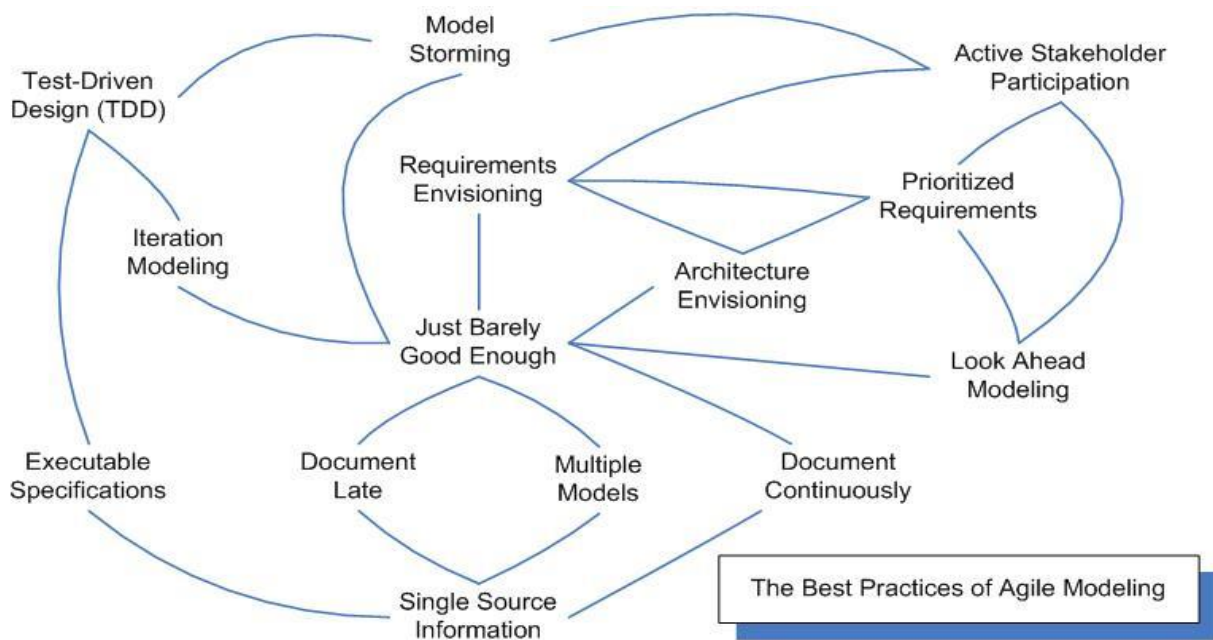
الشكل (٧) مراحل عمل دورة حياة تصميم وتطوير مشاريع FDD [35]

تقوم هذه المنهجية على مفهوم الميزة كما هو واضح، ويمكن تعريف الميزة "بأنها وظيفة صغيرة، ومهمة للزبون، ويمكن ترجمتها في الصيغة: <عمل><نتيجة><غرض>" [34]. يتم اختيار الميزات من توصيف المتطلبات الرسمية،

وذلك بحيث لا تتجاوز عملية تصميم وتطوير هذه الميزة أكثر من أسبوعين. تعرّف المنهجية FDD ستة أدوار أساسية وتوسع أدوار مساعدة. تستند جميع المخططات والرسوم البيانية والرموز في هذه المنهجية على لغة النمذجة UML، بالإضافة لمجموعة من الوثائق النصية.

٣-٧. Agile Modeling (AM)

النمذجة الرشيقية (AM) هي منهجية قائمة على الممارسات من أجل نمذجة وتوثيق النظم والبرمجيات بشكل فعال. "النمذجة الرشيقية (AM) هي نهج جديد لتنفيذ أنشطة النمذجة" [35]. "اقترحت النمذجة الرشيقية (AM) كمحاولة لتطبيق مبادئ XP و AD على أنشطة النمذجة والتصميم" [32]. كما هو الحال في المنهجية XP تقوم النمذجة الرشيقية على مجموعة من المبادئ، والقيم، والممارسات، من أجل تطوير برامج النمذجة بطريقة فعّالة وبسيطة. اشتقت معظم مبادئ وقيم النمذجة الرشيقية (AM) من المنهجية الرشيقية XP، وحددت سبع فئات من الممارسات في هذه المنهجية بناءً على المبادئ والقيم السابقة كما يبيّن الشكل (٨) التالي عدداً منها:



الشكل (٨) خريطة مجموعة الممارسات في المنهجية "النمذجة الرشيقية (AM)" [32]

تشكل أربع فئات من هذه الممارسات الفئات الأساسية للنمذجة الرشيقية (AM) وهي: (I) النمذجة المتزايدة والتكرارية. (II) العمل الجماعي. (III) البساطة. (IV) التحقق من الصحة. فيما تعتبر فئات الممارسات الثلاثة الأخرى اختيارية وهي: (I) الانتاجية. (II) التوثيق. (III) الحافز.

التركيز الرئيسي في هذه المنهجية على النمذجة والتوثيق، ولا تضمن أنشطة أخرى مثل: تطوير البرمجيات واختبارها، وإدارة المشاريع، وتطوير النظم وصيانتها. "يجب استخدام النمذجة الرشيقية (AM) مع إحدى المنهجيات الرشيقية الكاملة

الأخرى مثل: XP أو DSDM أو RUP [34]. لا تعرّف النمذجة الرشيقة (AM) أية أدوار للأفراد خلال مراحل العمل.

٤. مقارنة بين المنهجيات الرشيقة وملاءمتها لتطبيقات ذكاء الأعمال

إنّ المنهجيات الرشيقة قد قدمت حلولاً فعّالة في تطوير البرمجيات وتحسين عمليات نشرها وتسليمها، وتعتبر تطبيقات ذكاء الأعمال (BI) إحدى هذه المجالات التي يمكن فيها الاستفادة من هذه الحلول في مواجهة الصعوبات التي تم ذكرها سابقاً في هذا البحث.

نستعرض فيما يلي الجدول (٢)، للمقارنة بين هذه المنهجيات بناءً لمجموعة من المعايير، وهي: (I) السمات الأساسية. (II) الميزات الخاصة. (III) اكتمال النموذج. (IV) التوثيق. (V) حجم فريق العمل. (VI) تعريف أدوار فريق العمل. (VII) ملاءمتها لتطبيقات ذكاء الأعمال.

الجدول (٢) مقارنة بين المنهجيات الرشيقية وملاءمتها تطبيقات ذكاء الأعمال (BI)

المهجيّة	ASD	SCRUM	XP
المهجيّة	ASD	SCRUM	XP
السمات الأساسية	التكيف، العمل الجماعي التعاوني، التعلم، مقادة بالمهمة (mission-driven).	فرق عمل صغيرة، التكيف، الوضوح، الشفافية، دورات عمل دقيقة، العروض ومراجعة العمل (أسبوعاً) ٣٠-٧ دورة.	مقادة بالزبون (Customer-driven)، إطلاق نسخ متكررة، برمجة ثنائية، التركيز على الاختبار.
الميزات الخاصة	مراحل دورة حياة خطية غير متداخلة، يعتمد على المكونات.	مقابلات يومية قصيرة (١٥ بعد كل إضافة).	قصص المستخدم، إعادة الهيكلة، التطوير يبدأ الاختبار أولاً، سرعة المشروع.
النموذج اكتمال	مكتملة.	مكتملة.	مكتملة.
التوثيق	الكب والموقع الإلكتروني.	الكب والموقع الإلكتروني.	الكب والموقع الإلكتروني.
حجم فريق العمل	غير محدد.	١٠ أعضاء على الأكثر.	١٢ عضو على الأكثر.
تدوير أدوار فريق العمل	غير مفصلة.	مفصلة.	مفصلة.
ملاءمتها لذكاء الأعمال	تم اعتمادها من قبل (Alnoukari, 2012).	المنهجية المقترحة في بحثنا هذا.	غير مقترحة: تحتاج موارد خارجية (برمجة ثنائية)، يفضل لمشروع صغير، يحتاج دعم ممارسات إدارية.

Crystal family	DSDM	FDD	Agile Modeling
عائلة من الطرائق، متكيف مع مختلف حجوم المشاريع وتعقيدها.	مشاهدة للخصائص XP، بالإضافة لملاءمتها لأغراض العمل.	خمس خطوات أساسية للعمل، دورات قصيرة، تركز على المزايا.	يطبق مبادئ وممارسات المنهجيات الرشيقة على النمذجة.
فرق صغيرة، دورة كل ١-٣ أشهر، التواصل وجهاً لوجه.	استخدام النماذج الأولية، فرق عمل متعددة صغيرة (٢-٦ أشخاص).	الدمج بين المزايا والنمذجة الغرضية، توسعية.	التركيز الرئيسي على النمذجة والتوثيق.
غير مكتملة بعد.	مكتملة.	مكتملة.	غير مكتملة بعد.
الموقع الإلكتروني.	الكتب والموقع الإلكتروني.	الكتب والموقع الإلكتروني.	الكتب والموقع الإلكتروني.
حتى ٤٠ عضو في الفريق.	من ٢ وحتى ٦ أعضاء في كل فريق.	من ٥٠ وحتى ٢٥٠ عضو.	غير محدد.
مفصلة.	مفصلة.	مفصلة.	غير مفصلة.
غير مقترحة: غير مكتملة بعد.	غير مقترحة: يحتاج لفرق صغيرة متعددة، ينطبق على المشاريع الصغيرة، يقتصر الدعم للأعضاء المشتركين.	غير مقترحة: يحتاج دعم ممارسات إدارية، وعدد هائل لأعضاء فريق العمل.	غير مقترحة: غير مكتملة بعد ويحتاج لمنهجية أخرى للتكامل معها.

٥. لماذا المنهجية الرشيقية سكروم

تعتبر المنهجية سكروم (SCRUM) إحدى أهم هذه المنهجيات، فهي تؤمن إطار عمل متكيف ومتجاوب مع التغييرات والتطورات السريعة داخل المؤسسة نفسها من جهة والأسواق الخارجية من جهة أخرى. بناءً على المقارنة السابقة ودراسة هندسية المنهجية سكروم - التي يمكن مراجعتها في الملحق ٢ في نهاية هذا البحث - قمنا باعتماد المنهجية سكروم؛ لما تقدمه من خصائص ومميزات لمواجهة تحديات بيئات العمل المتطورة بشكل متسارع والتغييرات على مستوى المتطلبات والاحتياجات، ومن أهم هذه الخصائص:

- تعد المنهجية سكروم من أكثر المنهجيات الرشيقية انتشاراً واعتماداً من قبل الشركات حول العالم، لما تقوم عليه من ركائز ثلاث أساسية هي: (I) التكيف. (II) الشفافية. (III) الوضوح. وما تقدمه من أساليب لإدارة فرق العمل وتحديد أدوار أعضائها وصلاحياتهم بشكل مفصل.
- يعتبر التسليم المبكر والمستمر للمنتج إلى جانب المعلومات والبيانات الأولية القصوى في هذه المنهجية: ونحن في بيئة تطوير تطبيقات ذكاء الأعمال (BI) نحتاج لنشر المعلومات والتقارير بشكل مستمر ومتواصل خلال فترات زمنية تمتد بين أسبوعين وحتى شهرين كحد أقصى.
- تضمن المنهجية سكروم عملية بناء وتطوير تطبيقات ذكاء الأعمال (BI) بشكل قادرة على التجاوب مع المتغيرات الخارجية في بيئة العمل أو متطلبات الزبائن: فهي تقوم على فريق عمل متكامل مُدار ذاتياً يملك المهارات اللازمة لتطوير المنتج والقدرة على تحديد ما هو ضروري وما هو غير ضروري من جهة، ويملك الصلاحيات لتغيير جعبة كل شوط عند ورود أعمال أو متطلبات جديدة، وإعادة كتابة التقديرات اللازمة لهذا العمل.
- تتميز المنهجية سكروم بالبساطة والشفافية بين جميع الأفراد المشاركين بالعمل وأصحاب المصلحة، مما يُمكن تطوير تطبيقات ذكاء الأعمال (BI) وتوجيهها بحسب رغبة الزبائن والمستخدمين النهائيين: فالمنهجية تعرف مجموعة من المفاهيم التي يتشارك الجميع الفهم السليم والواضح لها، سواء أكانت جعبة الشوط أو جعبة المنتج ككل.
- تعرف المنهجية سكروم مجموعة من الأحداث الواضحة والمحددة الزمن بما يضمن عمل فرق تطوير تطبيقات ذكاء الأعمال (BI) في بيئة متغيرة بشكل مستمر ومتسارع: فتقوم على تجزئة العمل على مجموعة من الأشواط المتتالية تتضمن مجموعة من الأنشطة اليومية والأحداث في بداية ونهاية كل شوط بين أعضاء فريق التطوير من جهة وبين فريق التطوير وأصحاب المصلحة الآخرين من جهة أخرى، بما يسمح بعمل رجال الأعمال والمطورين معاً لتحديد أي قصور أو عقبات أو تغييرات في عملية التطوير، ويعطي القدرة لفرق التطوير لتحسين وضعها وزيادة فعاليتها لتجاوز هذه العقبات والتكيف مع المتغيرات الطارئة.

٦. خاتمة

إن أهمّ المشاكل والصعوبات التي تواجه عملية بناء تطبيقات ذكاء الأعمال (BI) وتطويرها، تعود للتغيير السريع في الشروط والأولويات والمتطلبات في بيئات العمل والأسواق. بالمقابل قدمت المنهجيات الرشيقة حلولاً لمواجهة التحديات في مجال تطوير البرمجيات، في ظل شروط الأسواق وبيئات العمل التي تتصف بعدم الاستقرار والتطور السريع. قمنا في هذا الفصل بتقديم لمحة عن بنية تطبيقات ذكاء الأعمال (BI) وأدواته، كما استعرضنا أهمّ المنهجيات الرشيقة، وأهم مميزات كل منها، وقمنا بالمقارنة بينها بناءً لخصائصها، ومميزاتها، وملاءمتها لتطبيقات ذكاء الأعمال، محاولين توضيح سبب اختيار المنهجية سكروم في حلّنا المقترح في هذا البحث، وذلك تمهيداً للفصول القادمة من الدراسة.

الفصل الثالث

الدّراسة المرجعيّة

١. مقدمة

بدأ الحديث حول نمذجة اكتشاف المعرفة في البيانات (KDD) في العام ١٩٨٩ خلال ورشة العمل (IJCAI-) اكتشاف المعرفة في البيانات، بداية من: (Piatetsky-Shapiro, 1994) و (Fayyad et al., 1996) الذين عرفوا الخطوات الأساسية لبناء تطبيقات التنقيب في البيانات وتحليلها، معرفين دورة حياة هذه التطبيقات. في جميع الأحوال تُعدّ المعرفة العامل المشترك بين جميع هذه النماذج كخروج نهائي لها.

نقوم في هذا الفصل بتحليل أبرز النماذج المعروفة في مجال اكتشاف المعرفة، والتي تختلف في مراحلها من حيث العدد، والدورات، والعمليات، والبنية، والمفاهيم والأدوات المطبقة، فسندم في البداية تصنيفاً لنماذج اكتشاف المعرفة الموجودة، ومن ثمّ نقدم إطاراً للمقارنة بين أبرز النماذج لتحليل مواضع الضعف والقوة فيها، لنختتم بمصفوفة السمات والخصائص المميزة لهذه الأعمال.

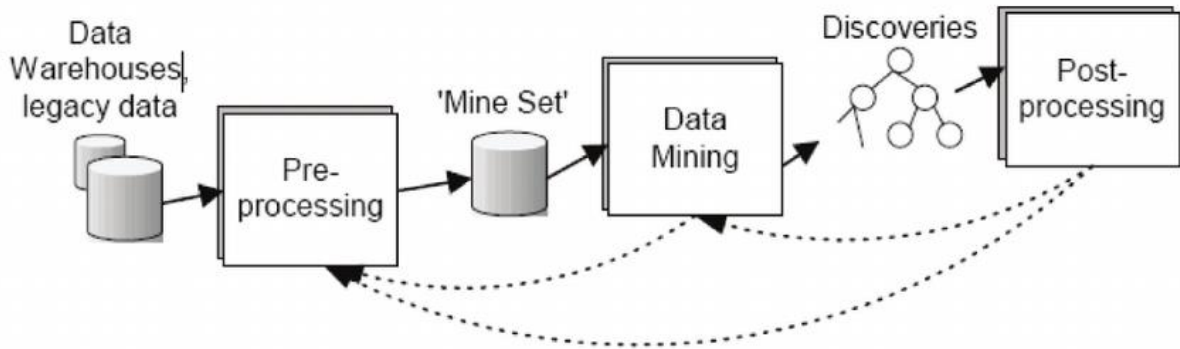
٢. تصنيف نماذج اكتشاف المعرفة

يمكننا تصنيف نماذج اكتشاف المعرفة في البيانات (KDD) إلى فئتين أساسيتين بحسب النهج الذي اتبعته، هما: (I) نماذج اكتشاف المعرفة التقليدية. (II) نماذج اكتشاف المعرفة الهجينة. سنقوم فيما يلي بعرض لمحة عن كلٍّ منها.

٢-١. نماذج اكتشاف المعرفة التقليدية

يعد هذا النهج الأكثر استخداماً في مجال اكتشاف المعرفة في البيانات (KDD)، بداية مع نموذج (Fayyad et al., 1996)، والكثير من النماذج الأخرى التي اعتمدت نفس الخطوات: فهم العمل، وفهم البيانات، ومعالجة البيانات، ونمذجة البيانات والتنقيب في البيانات، وتقييم النتائج، والنشر والإظهار. ظهرت فيما بعد العديد من النماذج المعتمدة على نموذج فياض مع إدخال عدد من التعديلات الهامة لتلافي نقاط الضعف فيه، كالنموذج (Feldens et al., 1998) الموضح بالشكل (٩)، والذي يقوم على المراحل الثلاثة الأساسية التالية: مرحلة ما قبل المعالجة، ومرحلة التنقيب في البيانات، ومرحلة ما بعد المعالجة. فيما بعد أضاف نموذج (Collier et al., 1998) مرحلة جديدة بعد مرحلة النشر هي مرحلة المراقبة، لمتابعة المنتج بعد نشره، وإضافة حلقات داخلية بين المراحل السابقة، لتتيح إمكانية التعديل، ومراعاة التغييرات بدلاً من المسار التقليدي المباشر. والجدير بالذكر هنا أنّ كلا النموذجين السابقين استخدم مستوى البيانات (DW) كمصدر للبيانات. أضافت النماذج اللاحقة مرحلتين أساسيتين هما: مرحلة فهم بيئة العمل (Business Understanding) التي تفيد في الفهم الكامل والشامل لبيئة العمل، ومرحلة فهم البيانات (Data Understanding) التي تساعد في معالجة البيانات بشكل أفضل.

في جميع الأحوال، إنّ معظم النماذج التي اتبعت النهج التقليديّ أغفلت دور الموارد البشرية والخبراء في مجال المعرفة والأعمال، الأمر الذي حثّ الباحثين في هذا المجال نحو إشراك العديد من الخبراء والمختصين في مجال التنقيب في البيانات واكتشاف المعرفة في النماذج الجديدة.

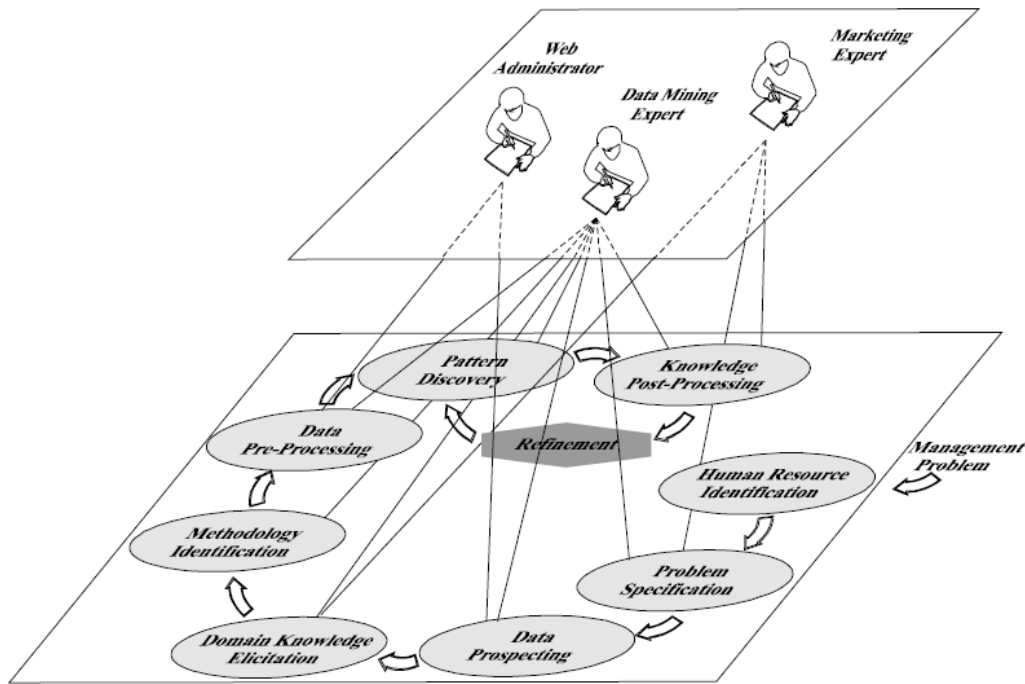


الشكل (٩) نموذج (Feldens et al., 1998) [36]

٢-٢. نماذج اكتشاف المعرفة الهجينة

حاول بعض الباحثين الاستفادة من بعض المفاهيم والأدوات، للتغلب على المشاكل والصعوبات التي تواجهها النماذج التقليدية، وتنظيم المعرفة المكتشفة، وإدارتها، ونقلها، وتخزينها. تطرق بعض الباحثين لأهمية اكتشاف المعرفة في

سجلات الويب، أو التنقيب في الويب (Web Mining)، فقدّم (Buchner et al., 1999) نموذج "Internet_enabled Knowledge Discovery" الموضح بالشكل (١٠) للاستفادة من بيانات الويب في عمليات التسويق والتجارة الإلكترونية، لجذب اهتمام الزبائن، والمحافظة عليهم، وزيادة المبيعات. وقدم (Pabarskaite and Raudys, 2007) مراجعة نقدية لعدد كبير من الأعمال في مجال التنقيب في الويب؛ تطرقوا خلالها لأهم القضايا المتداولة في هذا المجال، بما فيها: تنظيف البيانات، وتحديد المستخدمين، وتحليل جلساتهم.



الشكل (١٠) Internet_enabled Knowledge Discovery (١٩٩٩) [37]

بالمقابل عمل بعض الباحثين على الدمج بين المنهج التقليدي ومفاهيم هندسة الأنطولوجية، للاستفادة من هذه المفاهيم في تنظيم المعرفة المكتشفة. يعتبر Gotttroy من أشهر من عملوا في هذا النهج، الذي قدم نموذج (Ontology Driven Knowledge Discovery, 2010) في البحث الذي أعده لنيل درجة الدكتوراه في جامعة أوكلاهو للتكنولوجيا (AUT University)، محاولاً الاستفادة من هندسة الأنطولوجية في توصيف المعرفة، وتمثيلها، وإدارتها، ومكملة أدوات التنقيب في البيانات وخوارزمياتها، واكتشاف المعرفة دلاليًا. "يوجد حاليًا ثلاثة توجهات في هذا المنهج وهي: الأنطولوجية من أجل اكتشاف المعرفة في البيانات (Onto4KD)، واكتشاف المعرفة في البيانات (KDD4Onto) من أجل الأنطولوجية، والدمج بين التوجهين السابقين (Onto4KDD4Onto)" [38].

سعى آخرون لمواجهة الصعوبات والمشاكل في بيئة العمل، والتغيرات الديناميكية والمتسارعة على مستوى المتطلبات، والأولويات، والجدول الزمني، من خلال الاستفادة من خصائص المنهجيات الرشيقية (Agile)

(Methodologies)، التي ظهرت بداية القرن الحالي لمواجهة التحديات في مجال تطوير البرمجيات، وهي نفس التحديات التي تواجهها عملية اكتشاف المعرفة وبناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات. فظهرت العديد من الدراسات والتوصيات لأهمية الاستفادة من هذه المنهجيات في مجال اكتشاف المعرفة [9] [7].

يعتبر النموذج (ASD-DM, 2008) الذي قدمه Alnoukari من أوائل النماذج التي اعتمدت على المنهجيات الرشيقة في نمذجة اكتشاف المعرفة في البيانات (KDD)، ثم قام ببعض التعديلات عليه، وأعاد نشره باسم (ASD-BI, 2012).

٣. المعاملات التي بني عليها إطار المقارنة

نقدم هنا إطاراً للمقارنة بين الأعمال السابقة والمختارة التي سنناقشها لاحقاً بالتفصيل من حيث البنية، والمراحل، والأعمال، والمفاهيم، والأدوات التي تتضمنها، فإن تركيزنا ينصب في هذه المقارنة على تقييم نماذج اكتشاف المعرفة في البيانات المختارة في هذه الدراسة، وتحديد نقاط الضعف والقوة فيها، لتلافي نقاط الضعف فيها، والاستفادة من نقاط القوة فيها في الحل المقترح في هذا البحث. جرت المقارنة بناءً لمجموعة من المعاملات الأساسية وهي: (I) النهج المتبع. (II) الأنظمة التجارية الداعمة. (III) مخازن البيانات. (IV) طبيعة العمليات. (V) طبيعة الموارد البشرية. (VI) طبيعة المسارات. (VII) التكيف مع التغييرات. (VIII) البعد المعرفي. (IX) البعد الاستراتيجي. سنحاول فيما يلي توضيح كل منها:

٣-١. النهج المتبع

اتبعت معظم النماذج المقدمة لاكتشاف المعرفة في البيانات النهج التقليدي، فيما حاول بعض الباحثين الآخرين الاستفادة من مفاهيم وتقنيات أخرى في دعم عمليات تحليل البيانات لاستخلاص المعرفة وتنظيمها، نتم في هذا المعامل في تحديد النهج المتبع، والمفاهيم والأدوات المستخدمة في النماذج التي اتبعت نهجاً هجيناً.

٣-٢. نوع النموذج والأنظمة التجارية الداعمة

أعتمدت عدد من النماذج من قبل بعض الشركات والمؤسسات التجارية والصناعية، فأطلقت التطبيقات الخاصة بهذه النماذج ضمن منصة البرمجيات الخاصة بها. فيما بقيت النماذج الأخرى كنموذج بحثي أكاديمي. نتم في هذا المعامل بتحديد نوع نموذج اكتشاف المعرفة (صناعي أو أكاديمي)، والأنظمة التجارية الداعمة له.

٣-٣. مخازن البيانات

"يعد من أهمّ المعاملات في تقييم أيّ نمذجة لاكتشاف المعرفة في البيانات (KDD)، فهو يعرف الدّخل والخروج الأساسي لهذا النموذج" [39]. اعتمدت الكثير من النماذج على البيانات المجردة كمصدر للبيانات، وأغفلت أهمية الاستفادة من مستودع البيانات وأسواقها (DW\DMs). نتم في هذا المعامل بتحديد مصادر البيانات ووجهتها كمستودع البيانات وأسواقها (DW\DMs) ومخزن المعرفة والمعلومات (IKR).

٣-٤. طبيعة العمليات

يُتصد بالعمليات جميع المراحل والإجراءات التي يتضمنها نموذج اكتشاف المعرفة في البيانات (KDD) خلال دورة حياته. "وقد تمّ تعريف مجموعة من الخطوات، على أنّها خطوات أساسية ومهمة لبناء أي نموذج شامل لاكتشاف المعرفة في البيانات (KDD)، وهي: فهم بيئة العمل، وفهم البيانات، وتحديد الأهداف والفرضيات، واختيار عينات البيانات، وتحضير البيانات، والنمذجة أو التنقيب في البيانات، والتقييم، وأخيراً نشر النتائج" [10]. نتم في هذا المعامل بتحديد طبيعة العمليات التي تتضمنها دورة حياة نموذج اكتشاف المعرفة.

٣-٥. طبيعة الموارد البشرية

نتم في هذا المعامل بتحديد طبيعة الموارد البشرية التي يتضمنها نموذج اكتشاف المعرفة في البيانات (KDD). فيجب أن "يجوز نموذج اكتشاف المعرفة في البيانات (KDD) عدد من الأشخاص بملكون خبرات متنوعة بما في ذلك: محلل بيئة العمل، ومهندس البيانات، والمنقب في البيانات، وخبير في مجال العمل، ومهندس المعرفة، وأخيراً مدير استراتيجي" [10].

٣-٦. طبيعة المسارات

"يجب أن تحوي دورة حياة النموذج القدرة على العودة بمسارات داخلية إلى جانب الحلقات الخارجية" [10]. فقد نظرت للعودة لعملية سابقة خلال دورة حياة النموذج دون حدوث أي تغييرات على مستوى العمليات الأخرى. اعتمدت نماذج اكتشاف المعرفة (KDD Model) مجموعة من الخيارات على مستوى المسارات التي تتضمنها دورة حياتها، من أهمها المسارات التالية: (I) دورات أو حلقات خارجية. (II) مسارات داخلية. (III) تدفق ثنائي الاتجاه (Bi-directional Flow). (IV) تخطي إلى الأمام (forward skipping). (V) تدفق تسلسلي متابعي (Sequential Flow). نتم في هذا المعامل بتحديد طبيعة المسارات التي تحويها دورة حياة كل نموذج.

٣-٧. التكيف مع التغييرات

"يعتبر هذا المعامل أساسياً لبناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، حيث المتطلبات متغيرة باستمرار" [7]. نتم في هذا المعامل بتحديد كيف يمكن لنموذج اكتشاف المعرفة في البيانات (KDD) أن يكون متكيفاً مع التغييرات في بيئة العمل والمتطلبات.

٣-٨. البعد المعرفي

نتم في هذا المعامل بتحديد كيف يمكن لنموذج اكتشاف المعرفة في البيانات (KDD) أن يدعم عملية الحصول على البيانات ومشاركتها، "ففي النهاية المعرفة هي الخرج الأساسي لأي نموذج لاكتشاف المعرفة في البيانات (KDD)" [7]. فيتم تقييم نموذج اكتشاف المعرفة هنا من حيث الإمكانيات التي يقدمها للحصول على معرفة جديدة، ومن ثم طريقة تخزين هذه المعرفة ومشاركتها للاستخدام لاحقاً في المستقبل.

٣-٩. البعد الاستراتيجي

نتم في هذا المعامل بتحديد كيف يمكن لنموذج اكتشاف المعرفة في البيانات (KDD) المساعدة في صياغة أهداف المنظمات والشركات، ووضع خططها الاستراتيجية بعيدة المدى، والمساعدة في تحقيق هذه الاستراتيجيات والوصول إليها.

٤. دراسة الأعمال المختارة

سنناقش في هذا الفصل أهم النماذج السابقة لاكتشاف المعرفة في البيانات (KDD)، موضحين تفاصيل بنية كل نموذج، والمراحل، والعمليات، والمسارات خلال دورة حياته، وطبيعة الموارد البشرية في كل نموذج، وآليات تحصيل المعرفة، وإظهارها، وتخزينها. وصولاً لتحديد نقاط القوة فيها للاستفادة منها في حلنا المقترح، وتحديد نقاط الضعف فيها والتي سنعمل على تلافيها في هذا الحل. وهذه الأعمال المختارة هي الأعمال الموضحة بالجدول (٣) التالي:

الجدول (٣) نماذج اكتشاف المعرفة السابقة المختارة في دراستنا المرجعية

عنوان النموذج	اسم النموذج	الباحث أو الشركة	عام النشر	نوع النموذج	التهج المتبع
اكتشاف المعرفة في قواعد البيانات [2]	KDD	Fayyad et al.	١٩٩٦	أكاديمي	تقليدي

تقليديّ	صناعيّ	١٩٩٧	SAS Institute	SEMMA	[40] SEMMA
تقليديّ	صناعيّ	٢٠٠٨	CRISP-DM	CRISP-DM	Cross-Industry-Standard Process [41] for Data Mining
تقليديّ	أكاديميّ	٢٠٠٩	Markus Hofmann	DMLC	دورة حياة التنقيب في البيانات العامة. [10]
هجين	أكاديميّ	٢٠١٠	Paulo Gottgroy	ODKD	اكتشاف المعرفة المقادة بالأنطولوجية. [38]
هجين	أكاديميّ	٢٠١٢	Mouhib Alnoukari	ASD-BI	Adaptive Software Development- [7] Business Intelligence

تم اختيار هذه النماذج الستة السابقة بناءً على منهجياتها الإبداعية المتبعة، وتطبيقاتها في المجالين العمليّ والصناعيّ، والمفاهيم والأدوات التي حاولت إدخالها ودمجها مع اكتشاف المعرفة في البيانات (KDD)، فالنموذج (KDD, 1996) يعتبر من أوائل نماذج اكتشاف المعرفة، وقد اعتمدت عليه عدد كبير من النماذج التقليدية النهج اللاحقة، أما النموذج (DMLC, 2009) يعد من أكثر النماذج التقليدية تكاملاً في بنيتها على مستوى العمليات، والموارد البشرية المضمنة، وتعريف مصادر البيانات ووجهتها، بالمقابل إنّ النموذجين (SEMMA, 1997; CRISP-DM, 2008) من أشهر النماذج الصناعية وأكثرها استخداماً وفق عدد من الإحصائيات -قمنا في الملحق ١ بعرض إحداها-، أما النموذجين (ODKD, 2010; ASD-BI, 2012) فقد تم اختيارهما بناءً للأفكار والمفاهيم الجديدة التي حاولت إدخالها ودمجها في نموذجيهما؛ فيعتبر Gottgroy من أوائل من عمل على إدخال مفاهيم هندسة الأنطولوجية لاكتشاف المعرفة في البيانات بهدف تنظيم المعارف المكتشفة، بالمقابل كان Alnoukari من أوائل من قدموا نموذجاً واضحاً لاكتشاف المعرفة في البيانات يعتمد المنهجيات الرشيقية، للمساعدة في بناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات في بيئات عمل غير مستقرة ومتطورة باستمرار.

سنورد ملحقاً في نهاية هذا العمل يحوي العديد من نماذج اكتشاف المعرفة، والتي تمّ استبعادها من هذا النقاش كونها أقلّ أثراً من جهة، أو أنّها تعتمد على إحدى المنهجيات التي سنناقشها، ولكنها أقلّ إسهاماً من جهة أخرى، مع العلم أنّنا سنحاول قدر المستطاع الإشارة لهذه النماذج خلال مناقشة المنهجيات المعتمدة عليها وتبيان إسهاماتها الإضافية إن وجدت، واختلافها عن هذه المنهجيات.

سنقوم فيما يلي بعرض الأعمال المختارة محاولين توضيح بنية كل منها، وأبرز نقاط الضعف والقوة، لنلخص

سمات كل منها ضمن إطار المقارنة الذي قمنا ببناءه سابقاً:

٤-١. اكتشاف المعرفة في قواعد البيانات "KDD" (Fayyad et al, 1996)

يتألف نموذج قياض لنمذجة اكتشاف المعرفة في البيانات (KDD) من الخطوات الخمسة التالية، والموضحة بالشكل (١١):

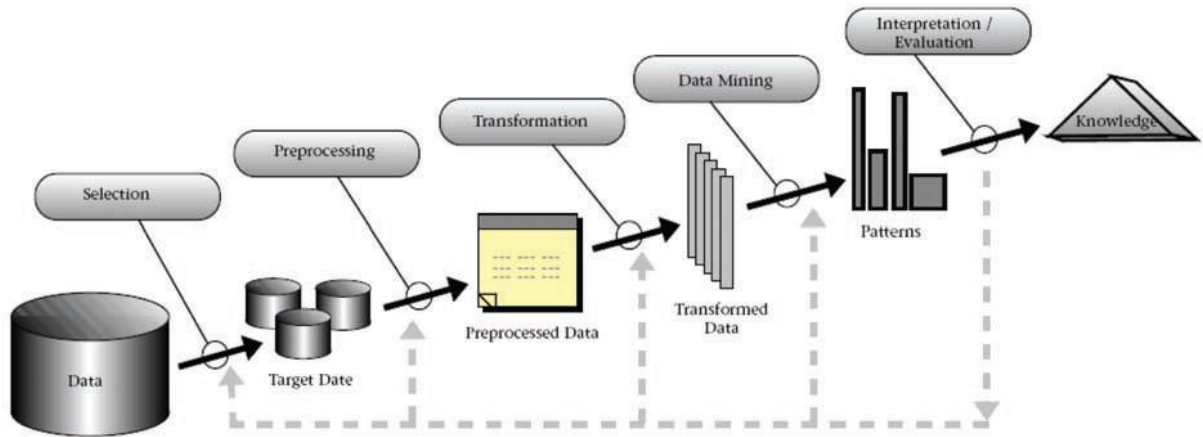
١. اختيار البيانات: بدورها تتألف من خطوتين: الأولى فهم مجال التطبيق والثانية تحديد العينة المستهدفة من البيانات من خلال مجموعة من الاستعلامات لاختيار البيانات المتعلقة بعملية الاكتشاف. الخرج النهائي لهذه المرحلة هو البيانات المستهدفة في اكتشاف المعرفة في البيانات (KDD).

٢. المعالجة الاستباقية للبيانات: يتم في هذه المرحلة معالجة بعض المشاكل مثل نقص البيانات وفلترتها الخاطئة، بما يضمن نتائج صحيحة لاكتشاف المعرفة في البيانات (KDD). الخرج النهائي لهذه المرحلة هو بيانات نظيفة.

٣. تحويل البيانات: وهي المرحلة الأخيرة من مراحل معالجة البيانات قبل تطبيق عمليات التحليل عليها. في هذه المرحلة يتم البحث عن الخصائص المميزة في البيانات من خلال تطبيق عمليات إنقاص الأبعاد والتحويل، لإيجاد تمثيل واضح وثابت للبيانات. الخرج النهائي لهذه المرحلة هو البيانات التي تم تحويلها.

٤. التنقيب في البيانات: تتألف هذه المرحلة من ثلاث خطوات: أولاً تحديد وظائف التنقيب في البيانات على ضوء الأهداف التي تم تحديدها سابقاً لاختيار طرق التنقيب في البيانات المتبعة من تصنيف وتجميع وعقدة وغيرها. ثانياً يتم اختيار خوارزميات التنقيب في البيانات والوسائل والخصائص اللازمة لتحديد الأنماط في البيانات. أخيراً تطبيق خوارزميات التنقيب في البيانات التي تم اختيارها لإيجاد نماذج البيانات وعرضها في صيغة تمثيل معينة. الخرج النهائي لهذه المرحلة هو مجموعة من النماذج والأنماط.

٥. التفسير والتقييم: تتألف هذه المرحلة من خطوتين اثنتين: أولاً تفسير وترجمة الأنماط التي تم تنقيبها ويمكن هنا العودة لأي خطوة سابقة والقيام بدورات أخرى، ليتم في نهاية هذه الخطوة عرض وإظهار النماذج والأنماط المستخرجة بالإضافة للبيانات المستخرجة من هذه النماذج والأنماط. ثانياً دعم المعرفة المستكشفة من خلال نظام أداء معين أو ببساطة من خلال توثيق النتائج وتوليد التقارير للأطراف التي ترغب بها. كما يتم في هذه المرحلة فحص وتصحيح التعارضات المحتملة في المعرفة المستخرجة مع المعارف الموثقة سابقاً. الخرج النهائي لهذه المرحلة هو المعرفة المحتملة.



الشكل (١١) نموذج Fayyad et al. ١٩٩٦ [2]

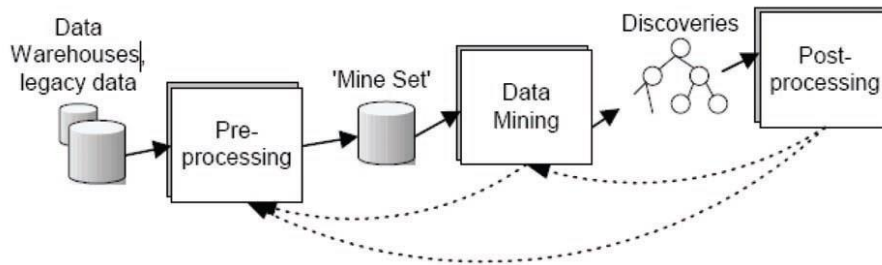
"يعدُّ هذا النموذج واحداً من أوائل المحاولات لنمذجة اكتشاف المعرفة في البيانات (KDD) في نموذج واحد متكامل" [42]. والأكثر اعتماداً من قبل النماذج اللاحقة في هذا المجال. ولكنّه عانى من بعض نقاط الضعف أهمها: اعتماده على البيانات المجردة على الرّغم من اهتمامه بعمليات اختيار البيانات، ومعالجتها، وتحويلها، وأغفل أهمية الاستفادة من مستودع البيانات وأسواقها (DW\DMs)، إلى جانب ذلك لم يوضّح خطوات نشر النتائج، أو تخزينها. يمكن تلخيص أبرز سمات هذا النموذج ضمن إطار المقارنة المعرف في هذه الدّراسة من خلال الجدول (٤) التالي:

الجدول (٤) سمات النموذج (KDD, 1996) ضمن إطار المقارنة المعرف

سمات النموذج	معامل المقارنة
اتبع نهجاً تقليدياً، ويعدّ حجر الأساس لمعظم النماذج التقليدية اللاحقة.	النهج المتبع
نموذج أكاديمي، ولكن الشركة الأميركية "Silicon Graphics, Inc." المتخصصة بمجال أجهزة الكمبيوتر وبرمجياتها قدمت برمجية خاصة في مجال التنقيب في البيانات تدعم نموذج (Fayyad et al., 1996)، تحت اسم "MineSet™".	نوع النموذج والأنظمة التجارية الداعمة
تشكل البيانات الجامدة الدّخل الأساسي للنموذج، دون الاستفادة من مستودع البيانات وأسواقها (DW\DMs).	مخازن البيانات
يتكوّن من خمس عمليات أساسية وهي: (I) اختيار البيانات. (II) المعالجة الاستباقية للبيانات. (III) تحويل البيانات. (IV) التنقيب في البيانات. (V) التفسير والترجمة.	طبيعة العمليات
من الملاحظ أنّ العمليات تتمحور حول تحضير البيانات، ومعالجتها، وتحويلها. وأغفل أهمية دراسة بيئة العمل وتحديد الأهداف في بناء تطبيقات	

ذكاء الأعمال (BI) والتنقيب في البيانات بشكل ناجح. كما لا توجد مرحلة نشر واضحة لمشاركة المعارف المكتشفة.	
تجاهل النموذج دور الموارد البشرية تماماً في دورة حياته.	طبيعة الموارد البشرية
اتسمت دورة حياة النموذج بالطابع التسلسلي التتابعي؛ فلا تمر بحلقات خارجية تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. ولا يوجد أية مسارات داخلية حتى نهاية مرحلة التقييم، والتي قد تساعد في العودة لعملية سابقة دون الحاجة لتكرار عمليات لم تتغير أي قضايا متعلقة بها.	طبيعة المسارات
يجعل الطابع التسلسلي التتابعي لدورة حياة النموذج موضوع التكيف مع التغييرات أمراً صعباً. ولم يتم التطرق لهذه النقطة في توثيق النموذج.	التكيف مع التغييرات
لم يهتم النموذج بالمعارف المكتشفة، فلم يتطرق النموذج لآليات تخزين المعرفة المكتشفة، وإظهارها، ونقلها، ومشاركتها، للاستفادة منها لاحقاً.	البعد المعرفي
أغفل النموذج أهمية توجيه أهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات لتحقيق الخطط الاستراتيجية للشركات، فلم تحوي بنية النموذج على مرحلة واضحة لوضع أهداف هذه التطبيقات، أو تحديد الفرضيات والقيود المتعلقة بيئة العمل والتي تم إغفالها في بنية النموذج.	البعد الاستراتيجي

من أهم النماذج اللاحقة التي اعتمدت على هذا النموذج في بنيتها: (I) النموذج (Feldens et al., 1998): الذي اقترح بنية مبسطة، والموضحة بالشكل (١٢)، تتألف هذه البنية من المراحل الثلاث التالية: المعالجة الأولية، والتنقيب في البيانات، والمعالجة اللاحقة. من أهم الإسهامات في هذا النموذج اعتماد مستودع البيانات (DW) كمصدر للبيانات، وإضافة حلقات داخلية ضمن دورة حياة النموذج للعودة لعمليات سابقة.

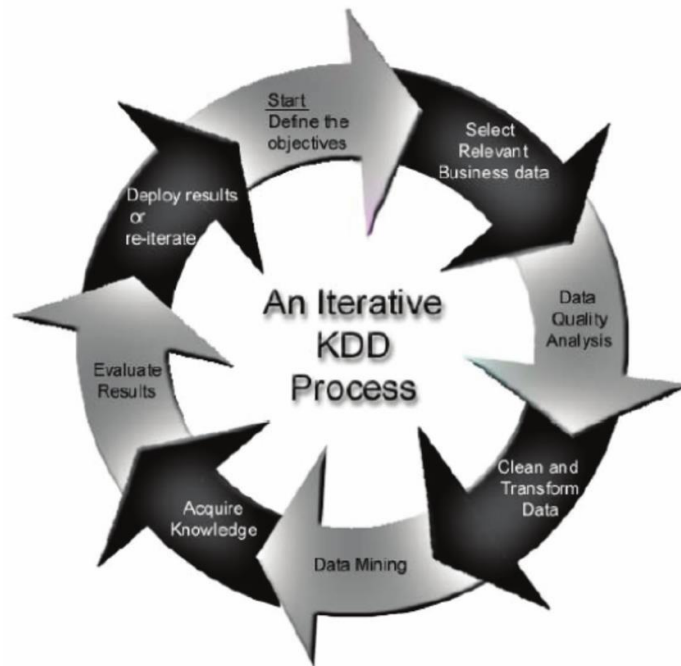


الشكل (١٢) نموذج (Feldens et al, 1998) [36]

(II) النموذج (Refined KDD Paradigm by Collier et al. 1998) [43]: الذي أضاف مرحلتين هامتين لبنيتها الموضحة بالشكل (١٣)، المرحتين هما: تعريف الأهداف، ونشر النتائج. من أهم الإسهامات الأخرى في

هذا النموذج دورة حياة النموذج التكرارية. ولكن يعاب على هذا النموذج عدم توضيح مصادر البيانات، وآليات تخزين المعارف المكتشفة، وتجاهله لوجود حلقات داخلية خلال دورة حياة النموذج.

أضافت العديد من النماذج اللاحقة العديد من الخطوات من أهمها: (I) فهم العمل (Business Understanding) التي تفيد في الفهم الكامل والشامل لبيئة العمل. (II) فهم البيانات (Data Understanding) التي تساعد في معالجة البيانات بشكل أفضل.



الشكل (١٣) نموذج (Refined KDD Paradigm by Collier et al., 1998) [43]

٤-٢. "SEMMA" by SAS Institute (1997)

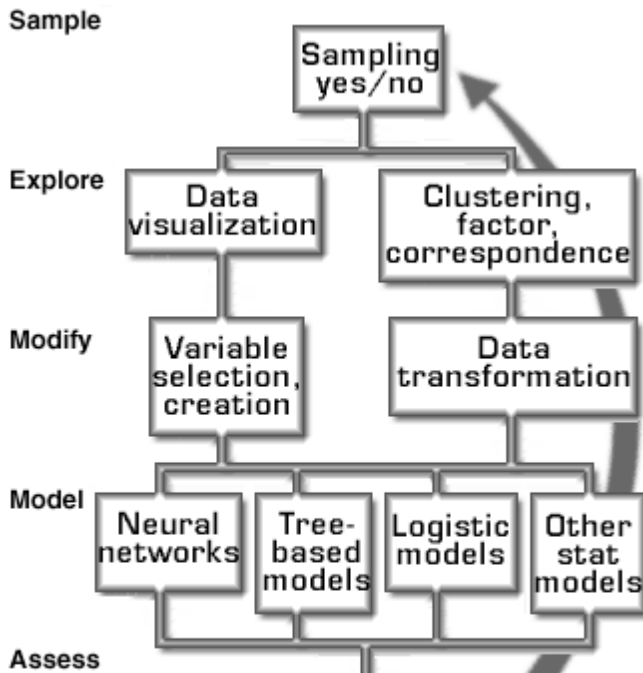
صُمم هذا النموذج من قبل الشركة الأمريكية "SAS Institute Inc."، وتم إدراجه ضمن منصة برمجيات لاكتشاف المعرفة من البيانات باسم "SAS Enterprise MinerTM". يتألف نموذج SEMMA من الخطوات الخمسة التالية الموضحة بالشكل (١٤):

١. اختيار عينات البيانات: يتم في هذه المحلة أخذ عينات من البيانات، يجب أن تكون هذه العينات كبيرة كفاية لتحتوي معلومات ذات قيمة، وصغيرة كفاية لتتم معالجتها سريعاً.
٢. الاستكشاف: يتم في هذه المرحلة البحث عن العلاقات، والاتجاهات، والشذوذ في البيانات للحصول على المعرفة.

٣. **التعديل:** يتمّ في هذه المرحلة اختيار النموذج من خلال تعديل البيانات بإنشاء المتغيرات الخاصة بها واختيارها وتحويلها.
٤. **النمذجة:** يتم في هذه المرحلة نمذجة البيانات بالشكل الذي يسمح للبرمجيات بالبحث عن تركيبة من البيانات وأنماطها للوصول للنتائج المرجوة بشكل موثوق.
٥. **التقييم:** يتم في هذه المرحلة تقييم صحة نتائج النمذجة وموثوقيتها.

"يمكن النظر لنموذج (SEMMA, 1997) على أنه التطبيق العملي لنموذج (Fayyad et al., 1996)" [44]. بإجراء مقارنة بسيطة بين العمليات والخطوات لكل من النموذجين يمكننا ببساطة التحقق من أنّهما متكافئين، إلّا أنّ ربط خطوات هذا النموذج بالتطبيقات البرمجية "SAS Enterprise MinerTM" جعله يصبح واحداً من أكثر المنهجيات استخداماً في مجال تطبيقات التنقيب في البيانات. يتشارك هذا النموذج معظم نقاط الضعف والقوة نفسها للنموذج (Fayyad et al., 1996) الذي قمنا بدراسته سابقاً، ولكنه بالمقابل تميّز عنه بمجموعة من النقاط أهمها: أشار لأهمية الاستفادة من مستودع البيانات وأسواقها (DW\DMS) من جهة، وأبرز أهمية مرور دورة حياة النموذج بعدة دورات تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة، ويمكن تلخيص أبرز سمات هذا النموذج ضمن

إطار المقارنة المعرف في هذه الدّراسة من خلال
الجدول (٥) التالي:



الشكل (١٤) خطوات عمل النموذج (SEMMA, 1997) [40]

الجدول (٥) سمات النموذج (SEMMA, 1997) ضمن إطار المقارنة المعرف

معامل المقارنة	سمات النموذج
النهج المتبع	اتبع نهجاً تقليدياً، فيعتبر التطبيق العملي لنموذج (Fayyad et al., 1996) كما ذكرنا.
نوع النموذج والأنظمة التجارية الداعمة	نموذج صناعي، تم تطويره من قبل الشركة الأمريكية "SAS Institute Inc."، وأطلقت الحلول والبرمجيات الخاصة به تحت اسم "SAS Enterprise Miner TM ".
مخازن البيانات	تشكل البيانات الجامدة الدخول الأساسي للنموذج، دون الاستفادة من مستودع البيانات وأسواقها (DW\DMs).
طبيعة العمليات	يتكوّن من خمس عمليات أساسية وهي: (I) اختيار عينات البيانات. (II) الاستكشاف. (III) التعديل. (IV) النمذجة. (V) التقييم. كما في النموذج (Fayyad et al., 1996) تتمحور العمليات حول تحضير البيانات، واختيار العينات، ومعالجتها، وتحويلها. واعتبر أنّ دورة حياة النموذج تقتصر على عمليات التنقيب في البيانات وتحليلها، أمّا الوظائف الأخرى كدراسة بيئة العمل، وتحديد أهداف تطبيقات التنقيب في البيانات، ونشر النتائج، تم قبل عمليات بناء تطبيقات التنقيب في البيانات وبعدها.
طبيعة الموارد البشرية	تجاهل النموذج دور الموارد البشرية تماماً في دورة حياته.
طبيعة المسارات	اتسمت دورة حياة النموذج بالطابع التسلسلي التتابعي، مع إمكانية العودة بحلقات خارجية تكرارية تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. بالمقابل لم يتم الحديث عن الحاجة لوجود مسارات داخلية، والتي قد تساعد في العودة لعملية سابقة دون الحاجة لتكرار عمليات لم تتغير أي قضايا متعلقة بها.
التكيف مع التغييرات	يجعل الطابع التسلسلي التتابعي لدورة حياة النموذج موضوع التكيف مع التغييرات أمراً صعباً. ولم يتم التطرق لأهمية بناء تطبيقات التنقيب في البيانات بشكل متكيف مع التغييرات الخارجية في توثيق النموذج.
البعد المعرفي	لم يهتم النموذج بالمعارف المكتشفة، فلم يتطرق النموذج لآليات تخزين المعرفة المكتشفة، وإظهارها، ونقلها، ومشاركتها، للاستفادة منها لاحقاً.
البعد الاستراتيجي	أغفل النموذج أهمية توجيه أهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات لتحقيق الخطط الاستراتيجية للشركات، فلم تحوي بنية النموذج

على مرحلة واضحة لوضع أهداف هذه التطبيقات، أو تحديد الفرضيات والقيود المتعلقة بيئة العمل والتي تمّ إغفالها في بنية النموذج.
--

٣-٤. Cross-Industry-Standard Process for Data Mining "CRISP-DM" by CRISP-DM (2008)

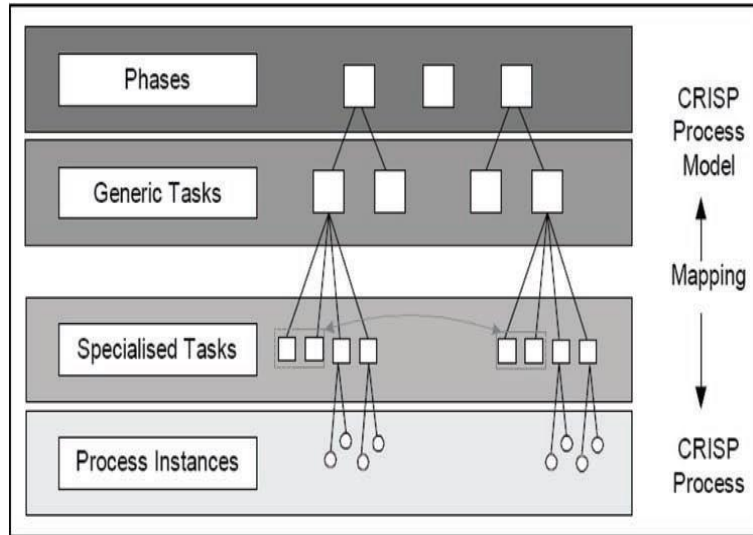
يعد هذا النموذج من أوائل نماذج اكتشاف المعرفة في البيانات (KDD) في المجال الصناعي، وأكثرها انتشاراً منذ تصميمه في العام ١٩٩٦ وفق معظم الإحصائيات، بما فيها إحصائيات موقع الويب (KDNuggets.com) الذي يشرف عليه Gregory Piatetsky-Shapiro مؤسس مفهوم اكتشاف المعرفة في البيانات إن أمكننا القول. تم تطوير النموذج عام ١٩٩٧ كأحد مشاريع الاتحاد الأوروبي بتمويل من مبادرة "ESPRIT". وقد تم تطويره من قبل مجموعة كبيرة من الشركات الأوروبية من أهمها الشركات الخمسة التالية: (I) SPSS. (II) Teradata. (III) Daimler (AG). (IV) NCR Corporation. (V) OHRA.

عُرضت النسخة الأولى من هذا النموذج خلال ورشة العمل "4th CRISP-DM SIG Workshop"، ليتم لاحقاً إطلاق إصدار الأول من CRISP-DM في العام ٢٠٠٠ كدليل لتطبيقات التنقيب في البيانات، وكان مدعوماً بشكل كبير من قبل مجموعة تدعى "CRISP-DM SIG"، التي توفّر إدخالاً من أكثر من ٢٠٠ مستخدم، وأدوات للتنقيب في البيانات، ومزودتي الخدمات. ليتم لاحقاً إطلاق الإصدار الثاني "CRISP-DM 2.0" في العام ٢٠٠٨، وتصبح شركة IBM الشركة الرئيسية التي تحتضن هذا النموذج، وأدرج هذا المنتج ضمن التطبيق "SPSS Modeler".

وضعت بنية هرمية لتوصيف النموذج (CRISP-DM, 2008) موضحة بالشكل (١٥) تتألف من المستويات

المجردة التالية:

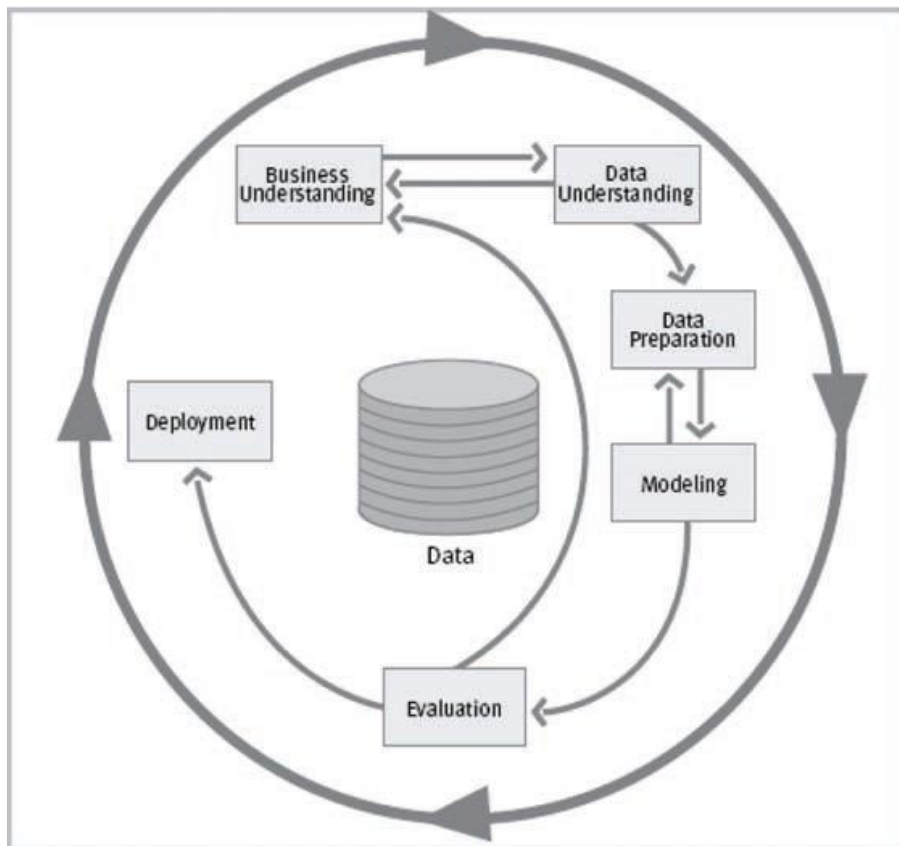
١. مراحل: يتكون كل نموذج من ستة مراحل كما سنرى لاحقاً في توصيفنا لبنية النموذج.
٢. مهام عامة: تتكون كل مرحلة من المراحل في المستوى الأعلى من مهام عامة تغطي جميع حالات التنقيب في البيانات بغرض أن تكون مستقرة وكاملة قدر الإمكان.
٣. مهام متخصصة: يتضمن توصيف الإجراءات اللازمة لتنفيذ المهام العامة في المستوى الأعلى.
٤. حالات عملية: يسجل مثل كل عملية الإجراءات والقرارات ونتائج استخراج البيانات الفعلية.



الشكل (١٥) البنية الهرمية لنموذج (CRISP-DM, 2008) [45]

كما أشرنا في توصيف البنية الهرمية لنموذج (CRISP-DM, 2008)، يتألف هذا النموذج من ستة مراحل

أساسية موضحة بالشكل (١٦) التالي:



الشكل (١٦) مراحل دورة حياة النموذج (CRISP-DM, 2008) [41]

سنقوم فيما يلي بعرض لمحة عن أهم المهام والأنشطة في هذه المراحل:

١. فهم العمل: يتم التركيز في هذه المرحلة على فهم متطلبات المشروع وأهدافه من منظور العمل، ليتم تحويلها فيما بعد إلى أهداف التنقيب في البيانات والخطة الاستباقية للمشروع. تتجزأ هذه المرحلة للخطوات التالية:

أ- تحديد أهداف العمل.

ب- تقييم الحالة.

ت- تحديد أهداف التنقيب في البيانات.

ث- توليد خطة المشروع.

٢. فهم البيانات: في هذه المرحلة يتم جمع البيانات، ومن ثم يتم القيام بالنشاطات اللازمة لجعل البيانات مألوفة للمستخدمين. تتكون هذه المرحلة من الخطوات التالية:

أ- جمع البيانات البدائية.

ب- توصيف البيانات.

ت- استكشاف البيانات.

ث- التحقق من جودة البيانات.

٣. تحضير البيانات: وتشتمل هذه المرحلة على جميع النشاطات اللازمة لبناء عينة البيانات النهائية التي سيتم استخدامها في المرحلة اللاحقة (النمذجة). تشتمل هذه المرحلة على الخطوات التالية:

أ- اختيار البيانات.

ب- تنظيف البيانات.

ت- هيكلية البيانات.

ث- تجميع البيانات.

ج- صياغة البيانات.

٤. النمذجة: يتم التركيز في هذه المرحلة على اختيار وتطبيق تقنيات النمذجة الملائمة. حيث يتم معايرة معايير تقنيات النمذجة التي للحصول على قيم مثالية، وهنا قد تتم العودة للخطوة السابقة بشكل متكرر للحصول على صيغة بيانات محددة من أجل تقنيات النمذجة. ويمكن تلخيص خطوات هذه المرحلة بالخطوات التالية:

أ- اختيار تقنيات النمذجة.

ب- توليد تصميم الاختبار.

ت- خلق النماذج.

ث- تقييم النماذج المولدة.

٥. التقييم: تعد هذه الخطوة حيوية بالنسبة للنموذج للتأكد من أن النماذج المختارة تلائم أهداف العمل، فإحدى أهم أهداف هذه المرحلة التأكد من عدم وجود أي قضية تخص العمل لم يتم أخذها بعين الاعتبار. تتألف مرحلة التقييم من الخطوات التالية:

أ- تقييم النتائج.

ب- مراجعة المعالجة.

ت- تحديد الخطوة التالية.

٦. النشر: تركز هذه الخطوة على تنظيم المعرفة المكتشفة وعرضها بطريقة تمكن من استخدامها، فهي نهاية دورة حياة عملية التنقيب في البيانات، ويمكن أن تكون بسيطة لدرجة توليد تقرير التنقيب في البيانات، أو صعبة لدرجة تحقيق عملية التنقيب في البيانات في المؤسسة أو الشركة. مرحلة النشر تنقسم للخطوات التالية:

أ- التخطيط للنشر.

ب- التخطيط للمراقبة والصيانة.

ت- توليد التقرير النهائي.

ث- مراجعة الخطوات الجزئية للعملية.

تمّ اعتماد نموذج (CRISP-DM) في العديد من مشاريع التنقيب في البيانات منذ إصدار النسخة الأولى في العام ٢٠٠٠، وذلك لدعمه من قبل عدد كبير من الشركات المهمة في أوروبا، بالإضافة للتوثيق الجيد لمراحل العمل، والأنشطة والمهام التي تتضمنها كل مرحلة. ولكن هذا النموذج كغيره من النماذج التقليدية عانى من نقاط ضعف أهمها: دخل النموذج البيانات المجردة، وأغفل أهمية الاستفادة من مستودع البيانات وأسواقها (DW\DMs)، كما لم يتطرق لآليات تخزين المعرفة المكتشفة، وإظهارها، ومشاركتها، بالإضافة لأغفاله أهمية الاستفادة من الموارد البشرية ودورها في أي مشروع قائم على المعرفة ومشاركتها. يمكن تلخيص أبرز سمات هذا النموذج بالجدول (٦) التالي:

الجدول (٦) سمات النموذج (CRISP-DM, 2008) ضمن إطار المقارنة المعرف

سمات النموذج	معامل المقارنة
اتباع نهجاً تقليدياً، ولكنه تميّز بتوثيق جيد لجميع المراحل وكافة المهام والأنشطة خلال هذه المراحل.	النهج المتبع
نموذج صناعي، تعتبر شركة IBM الشركة الرئيسية التي تحتضن هذا النموذج حالياً، وأدرج هذا المنتج ضمن برمجيات "SPSS Modeler".	نوع النموذج والأنظمة التجارية الداعمة

تشكل البيانات الجامدة الدّخل الأساسي للنموذج، دون الاستفادة من مستودع البيانات وأسواقها (DW\DMs).	مخازن البيانات
يتكوّن من ستة عمليات أساسية وهي: (I) فهم بيئة العمل. (II) فهم البيانات. (III) تحضير البيانات. (IV) النمذجة. (V) التقييم. (VI) النشر. يعتبر من أوائل النماذج التقليدية التي تضمن مرحلتين: فهم بيئة العمل وفهم البيانات بشكل صريح، وما تلعبه هاتين المرحلتين من دور هام في نجاح تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات. إلا أنه أغفل أهمية وجود عملية لتحديد أهداف هذه التطبيقات وتحديد الفرضيات والقيود، ما "يعطي رؤية أفضل لبيئة العمل، ويضيء على نقاط أخرى فيها" [22]. في جميع الأحوال يتميّز النموذج بتوثيق جيّد للمراحل.	طبيعة العمليات
تجاهل النموذج دور الموارد البشرية تماماً في دورة حياته.	طبيعة الموارد البشرية
اتسمت دورة حياة النموذج بالطابع التسلسلي التتابعي، فمن "العيوب الأساسية في هذا النموذج هي دورة حياته الخطية والمتابعة" [46]. على الرّغم من إمكانية العودة بحلقات خارجية تكرارية تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. بالمقابل لم يتم الحديث عن الحاجة لوجود مسارات داخلية حتى مرحلة التقييم، بحيث يتم العودة بدورة أخرى لتصحيح النتائج، على الرّغم من عدم حدوث أية تغييرات على مستوى عمليات أخرى كفهم بيئة العمل أو فهم البيانات.	طبيعة المسارات
يجعل الطابع التسلسلي التتابعي لدورة حياة النموذج موضوع التكيّف مع التغييرات أمراً صعباً. ولم يتم التطرق لأهمية بناء تطبيقات التنقيب في البيانات بشكل متكيف مع التغييرات الخارجية في توثيق النموذج.	التكيف مع التغييرات
لم يهتم النموذج بالمعارف المكتشفة، فلم يتطرق النموذج لآليات تخزين المعرفة المكتشفة، وإظهارها، ونقلها، ومشاركتها، للاستفادة منها لاحقاً.	البعد المعرفي
أغفل النموذج أهمية توجيه أهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات لتحقيق الخطط الاستراتيجية للشركات، فلم تحوي بنية النموذج على مرحلة واضحة لوضع أهداف هذه التطبيقات، أو تحديد الفرضيات والقيود، وإنما تاه بين أهداف العمل وهي مهمة يتم إنجازها خلال عملية فهم بيئة العمل، وأهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في المعطيات.	البعد الاستراتيجي

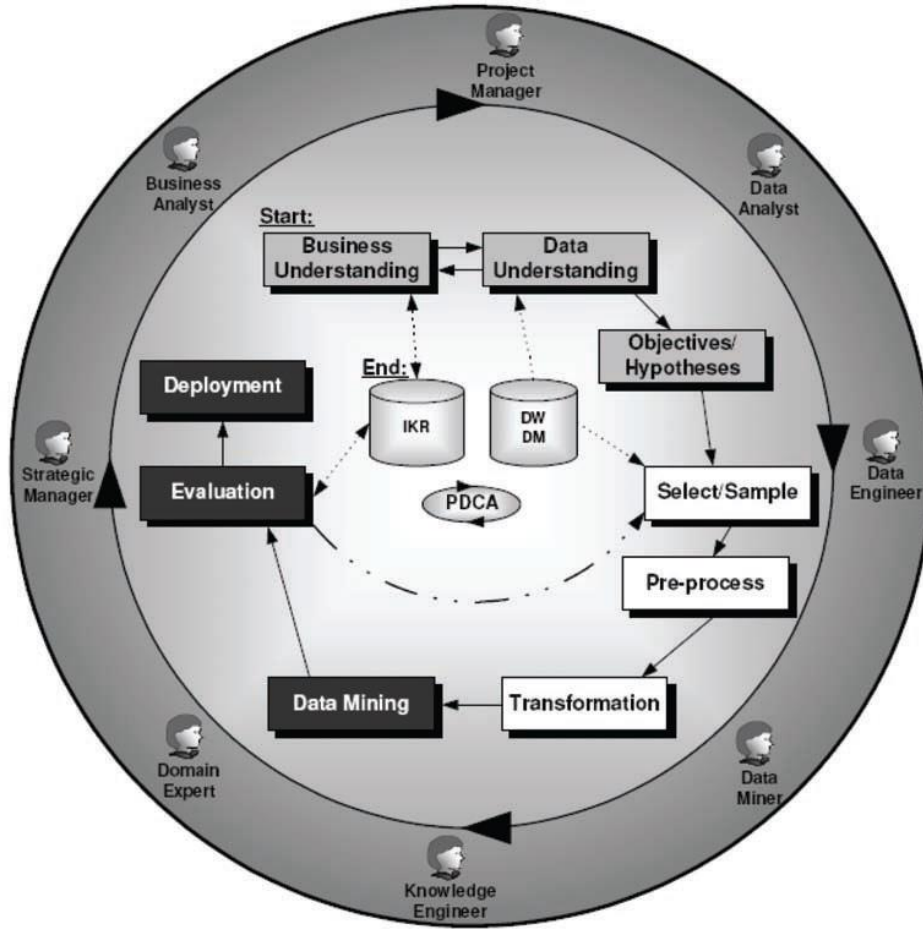
فأهداف التنقيب في المعطيات تحقق أهداف العمل في النهاية ولكن لا يمكن أن تكون هي نفسها.

٤-٤. دورة حياة التنقيب في البيانات العامة "DMLC" (Hofmann, 2009)

اقترح Hofmann هذا النموذج في العام ٢٠٠٣، وتمّ بناء المراحل والعمليات التي تتضمنها دورة حياة النموذج اعتماداً على النموذج (CRISP-DM, 2000) مع تجنب نقاط ضعف هذا النموذج والنماذج السابقة، وقام بإعادة نشره في العام ٢٠٠٩. تتألف دورة حياة النموذج DMLC من تسع خطوات - كما يظهر الشكل (١٧) - تمّ تجميعها في المراحل الثلاثة التالية:

١. **مرحلة تحضير الفرضيات والأهداف:** تتألف من ثلاث خطوات: فهم بيئة العمل، وفهم البيانات، وتعريف الأهداف والفرضيات.
٢. **مرحلة تحضير البيانات:** هذه المرحلة تقوم على الخطوات الثلاثة التالية: اختيار البيانات، والمعالجة الاستباقية للبيانات، وتحويل البيانات.
٣. **مرحلة الاكتشاف والتحقق:** خطوات هذه المرحلة: التنقيب في البيانات، والتقييم، والنشر.

تتكوّن مخازن البيانات في هذا النموذج من مستودعات البيانات وأسواقها (DW\DMS) بالإضافة لمخازن المعلومات والمعرفة (IKR). ويّدار هذا النموذج من قبل مجموعة من الموارد البشرية التي يجب تضمينها في دورة حياة مشروع التنقيب عن البيانات وهي: الخبير بمجال العمل، ومحلل البيانات، ومحلل بيئة العمل، ومهندس البيانات، والمنقب في البيانات، ومهندس المعرفة، والمدير الاستراتيجي. "تضاعف المشاركة الجيدة والكبيرة للموارد البشرية في مشاريع التنقيب في البيانات من صحة النتائج وجودتها" [7].



الشكل (١٧) مراحل النموذج دورة حياة التنقيب في البيانات "DMLC" المعتمدة من قبل (Hofmann and Tierney, 2007) [47]

أبرز ما يميز هذا النموذج عن النماذج التقليدية الأخرى بنيته المتكاملة على مستوى العمليات، والموارد البشرية، ومخازن البيانات. "استخدام الجوانب الثلاثة (العمليات، الموارد البشرية، مخازن البيانات) يعزز دور حياة مشاريع التنقيب في البيانات، ويزيد من شموليتها وتكاملها" [10]. فيعدّ من أوائل النماذج في مجال اكتشاف المعرفة في البيانات (KDD) التي أشارت لأهميّة مرحلة تحضير الأهداف والفرضيات، إلى جانب كل من عمليتي فهم بيئة العمل وفهم البيانات. "في هذا السياق يجب أن تتماشى الخطوات الثلاث السابقة جنباً إلى جنب حتى يتمّ تحديد الأهداف والفرضيات بشكل واضح ومحدّد" [10].

على الرّغم من ذلك إنّ هذا النموذج لا يزال يواجه بعض نقاط الضعف، من أهمها: عدد العمليات الكبير خلال دورة حياة النموذج، التي تتكون من ٩ عمليات البعض منها يقوم على عمل صغير جداً، كعملية المعالجة اللاحقة للبيانات وعملية التحويل، ففي الأولى يتم تنظيف عينات البيانات، وفي الثانية يتم إجراء عمليات تجميع (aggregation) وتقليص (reduction) على عينات البيانات، مما يجعل هذه العمليات أقرب لأن تكون إجرائية أو مهمة. بالإضافة للمشاركة

الضخمة للموارد البشرية والخبراء، "وكان النموذج بُني لمشاريع التنقيب في البيانات واسعة النطاق" [7]. فيمكن تلخيص أبرز سمات هذا النموذج ضمن إطار المقارنة المعرف في هذه الدّراسة من خلال الجدول (٧) التالي:

الجدول (٧) سمات النموذج (DMLC, 2009) ضمن إطار المقارنة المعرف

معامل المقارنة	سمات النموذج
التّهج المتبع	اتبع نهجاً تقليدياً، ويعدّ من أكثر النماذج تكاملاً في بنيتها وتوصيفها للعمليات والموارد البشرية ومخازن البيانات المضمنة في دور حياته.
نوع النموذج والأنظمة التجارية الدّاعمة	نموذج أكاديمي.
مخازن البيانات	تتكوّن مخازن البيانات في النموذج (DMLC, 2009) من: مستودع البيانات وأسواقها (DW\DMs)، ومخزن المعرفة والمعلومات (IKR). يشكل مستودع البيانات وأسواقها (DW\DMs) الدّخل الأساسي للنموذج، كما يتمّ تخزين البيانات المترفّعة (Metadata) الناتجة عن عملية فهم البيانات في هذا المخزن. فيما يشكل مخزن المعرفة والمعلومات (IKR) الخرج الأساسي للنموذج، ويتمّ تخزين المعارف المكتشفة ضمنه للاستفادة منها في دورات النموذج اللاحقة أو في مشاريع أخرى.
طبيعة العمليات	تتكوّن دورة حياة النموذج (DMLC, 2009) من تسع عمليات أساسية وهي: (I) فهم بيئة العمل. (II) فهم البيانات. (III) وضع الأهداف وتحديد الفرضيات. (IV) اختيار عينات البيانات. (V) المعالجة اللاحقة للبيانات. (VI) التحويل. (VII) التنقيب في البيانات. (VIII) التقييم. (IX) النشر. كما أسلفنا يعدّ نموذج (DMLC, 2009) من أوائل النماذج التي أشارت لأهمية عملية تحديد الأهداف ووضع الفرضيات في نجاح بناء تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، وتحقيق أهدافها. ولكنه عانى من عدد العمليات الكبير، فبعض العمليات تقوم على عمل صغير أقرب لأن يعبر عن مهمة أو نشاط من أن يعبر عن عملية ضمن دورة حياة النموذج. "يجب أن يكون عدد العمليات قابلاً للإدارة، ويضيف لتخطيط مشاريع التنقيب في البيانات وتنفيذها، بدلاً من أن يسبب إرباكاً أو ضياعاً خلال دورة حياة هذه المشاريع" [10].
طبيعة الموارد البشرية	يتضمن النموذج (DMLC, 2009) الموارد البشرية التالية: (I) مدير المشروع. (II) محلل البيانات. (III) مهندس البيانات. (IV) المنقب في

<p>البيانات. (V) مهندس المعرفة. (VI) خبير مجال العمل. (VII) المدير الاستراتيجي. (VIII) محلل الأعمال.</p> <p>يتميز النموذج بتوصيفه الجيد للأفراد والخبرات المضمنة في بنيته، وتوصيف مهمة كل منها ودورها خلال دورة حياته، ولكن يعاب عليه العدد الهائل لهذه الأفراد والخبرات، مما قد يشكل مشاكل وصعوبات في التعامل مع مشاريع ذكاء الأعمال (BI) والتنقيب في البيانات الصغيرة.</p>	
<p>اتسمت دورة حياة النموذج (DMLC, 2009) بخيارات متعددة على مستوى طبيعة المسارات؛ فتمر بحلقات خارجية تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. بالإضافة للمسارات الداخلية التي قد تساعد في العودة لعملية سابقة دون الحاجة لتكرار عمليات لم تتغير أي قضايا متعلقة بها. "المسارات الداخلية ضرورية للحصول على نتائج صحيحة لتطبيقات التنقيب في البيانات دون المرور عبر العمليات الأخرى التي تبقى على حالها" [10]. في جميع الأحوال وجود المسارات السابقة لم يؤثر على الحفاظ على الطابع التسلسلي التتابعي خلال دورة حياة النموذج بما يحقق الشمولية خلالها.</p>	طبيعة المسارات
<p>تمكن دورة حياة النموذج (DMLC, 2009) من الاستفادة من الخبرات المتراكمة خلال الدورات السابقة، وتصحيح الأخطاء بأقل جهد ممكن. ولكنها لا توضح آلية التعامل مع المتغيرات على مستوى المتطلبات أو أولويات العمل. ولم يتطرق Hofmann أبداً لهذه النقطة خلال توثيق النموذج.</p>	التكيف مع التغييرات
<p>يعتبر النموذج (DMLC, 2009) من أكثر النماذج التقليدية اهتماماً بآليات تنظيم المعرفة المكتشفة، وتخزينها، وإظهارها. فيتم تخزين المعرفة المكتشفة في مخزن المعرفة والمعلومات (IKR). ويسهم مهندس المعرفة المضمن في هذا النموذج في عمليات تحصيل المعارف والمعلومات المكتشفة، وهيكلتها بصيغة معيارية لتخزينها وإظهارها. "لا يقتصر دور مهندس المعرفة على ضمان تحصيل المعرفة، وإنما يلعب دوراً أساسياً في ضمان إظهارها وتخزينها بالشكل الذي يمكن من إعادة استخدامها" [48].</p>	البعد المعرفي
<p>أشار Hofmann في نمودجه لأهمية توجيه أهداف تطبيقات التنقيب في البيانات لتحقيق الخطط الاستراتيجية للشركات، فاحتوت بنية النموذج على مرحلة واضحة لوضع أهداف هذه التطبيقات، وتحديد الفرضيات والقيود</p>	البعد الاستراتيجي

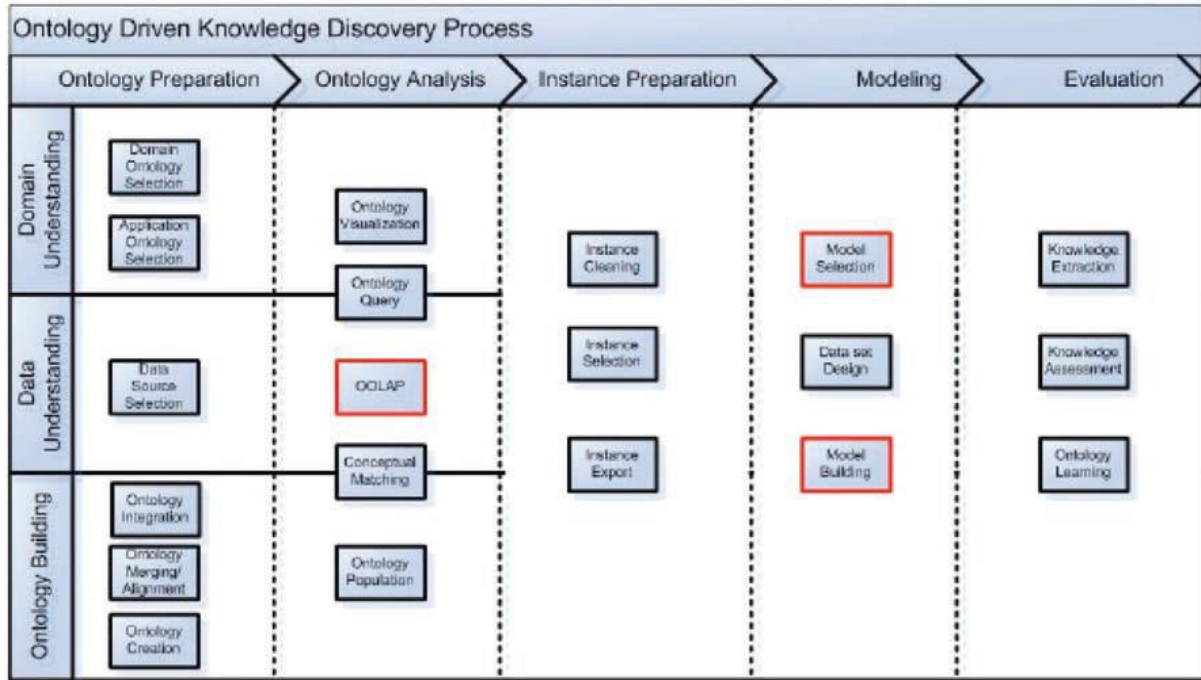
<p>المتعلقة ببيئة العمل. كما إنّ المدير الاستراتيجي المضمن في دورة حياة النموذج يلعب دوراً أساسياً في تطوير الأداء الاستراتيجي من خلال البيانات والدراسات حول بيئة العمل، وقيودها، وفرص المنافسة فيها.</p>
--

٤-٥. اكتشاف المعرفة المقادة بالأنطولوجية "ODKD" (Gottgortory, 2010)

"يعتبر هذا النموذج هو أحد أوائل المحاولات في الدمج بين المفاهيم في مجال المعرفة (كتعبير عن الأنطولوجية)، ومفاهيم اكتشاف المعرفة في البيانات (KDD) في دورة حياة واحدة صريحة وواضحة" [7].

أشارت العديد من الأبحاث في هذا المجال لأهمية الاستفادة من مفاهيم هندسة الأنطولوجية في نمذجة اكتشاف المعرفة في البيانات (KDD). من أهم هذه الأبحاث التي اقترحت منهجيات مقادة بالأنطولوجية للحصول على المعرفة، من هذه الأبحاث: (Yoon et al., 1999) [49] و (Phillips and Buchanan, 2001) [50] و (Gottgortroy et al, 2003) [51] و (Svatek et al, 2005) [52] و (Svatek et al, 2006) [53] وأخيراً (Kuo et al, 2007) [54].

يعتمد النموذج ODKD على كل من هندسة الأنطولوجية والمنهجية CRISP-DM. ويتألف من المراحل الخمسة التالية الموضحة في الشكل (١٨)، حيث تم تجزئة المرحلتين البدائيتين لثلاث قنوات: فهم مجال العمل، وفهم البيانات، وبناء الأنطولوجية. تهدف كل من قناتي فهم مجال العمل والبيانات إلى فهم المشكلة، وتحديد مصدر البيانات، فيما ترتبط قناة بناء الأنطولوجية بعمليات بناء نماذج الأنطولوجية وإظهارها. بالمقابل ترتبط مرحلتي تحضير النسخة والنمذجة بعمليات التنقيب في المعطيات، حيث يتم في مرحلة التحضير جميع الأنشطة اللازمة لاختيار البيانات وتنظيفها، بالشكل الذي تصبح فيه ملائمة لعمليات النمذجة والتنقيب في البيانات. أما المرحلة الأخيرة مرحلة التقييم، فيتم تقييم النتائج خلالها على أساس المعرفة السابقة، والنسخ الجديدة التي تم تضمينها في النموذج الأنطولوجي.



الشكل (١٨) مراحل ومهام نموذج اكتشاف المعرفة المقادة بالأنطولوجية (ODKD, 2010) [38]

سنقوم فيما يلي باستعراض مراحل النموذج (ODKD, 2010):

١. **تحضير الأنطولوجية:** وهي المرحلة البدائية التي يتم خلالها: فهم البيانات، وسياق العمل، وجمع المتطلبات، بالإضافة

لمهام تحضير البيانات، وتحديد أهداف اكتشاف المعرفة في البيانات (KDD). تم تجزئتها لثلاث قنوات:

a. فهم مجال العمل: تتألف من مهمتين اثنتين: مهمة اختيار أنطولوجية مجال العمل لتحديد المعرفة التي

تغطي منظور المشكلة، ومهمة اختيار أنطولوجية التطبيق التي يتم التعامل فيها مع: تحديد وتقييم

الأنطولوجية، وتحليل المعرفة، وخطة تمثيل المعرفة. "يتم تجميع أنطولوجيات طيبة عامة عند التعامل مع

مشكلة طيبة خلال اختيار أنطولوجية مجال العمل، فيما تهدف اختيار أنطولوجية التطبيق لاختيار نماذج

محددة تخص المشكلة من هذه الأنطولوجيات" [38].

b. فهم البيانات: الاهتمام هنا بأن تصبح البيانات التي تم جمعها بناءً على مهام جمع المتطلبات السابقة

واضحة ومألوفة، كما تغطي الخطوات اللازمة لتحديد مشكلات جودة البيانات وتمثيلها من خلال

الأنطولوجية المختارة.

c. بناء الأنطولوجية: آخر قناة في مرحلة تحضير الأنطولوجية، وتتألف من المهام الثلاثة التالية: (I) مكملة

الأنطولوجية التي تتعامل مع التكامل بين الأنطولوجية المختارة، وفهم بيئة العمل بهدف تحديد إطار

نطاق المشكلة. (II) محاذاة الأنطولوجية ودمجها بهدف تحديد المفاهيم والعلاقات بين أنطولوجيتين اثنتين

لصيغة مجموعة جديدة من العلاقات والمفاهيم التي توضح نطاق المشكلة. (III) بناء الأنطولوجية

بهدف الدمج بين المعرفة الخاصة بالمشكلة ومجموعة من المفاهيم والعلاقات التي تم توليدها في المهمة السابقة.

٢. تحليل الأنطولوجية: في هذه المرحلة تتم الخطوات الأولى لاكتشاف نموذج الأنطولوجية، وتنقسم هذه المرحلة للقنوات الأربع التالية: إظهار الأنطولوجية، والاستعلام عن الأنطولوجية، ومطابقة المفاهيم، وتحديث الأنطولوجية. حيث تعود خطوات الاستعلام وإظهار الأنطولوجية لاكتشاف نموذج الأنطولوجية باستخدام وسائل البحث والتصور، بالمقابل تتعامل مطابقة المفاهيم وتحديث الأنطولوجية مع بناء قاعدة المعرفة.

٣. تحضير النسخة: هذه المرحلة تحوي جميع الأنشطة المتعلقة بميكلة عينة البيانات المثالية التي ستغذي مرحلة النمذجة لاحقاً. تحوي هذه المرحلة أربع قنوات:

a. تنظيف النسخة: كما في عملية تنظيف البيانات يتم هنا إزالة الأخطاء وإزالة التضارب في النسخة لتحسين جودة هذه النسخة التي ستغذي عملية نمذجة الأنطولوجية.

b. اختيار النسخة: وتركز هذه المهمة على تقديم معنى للبيانات والحد من المعاملات من خلال فهم المجال.

c. تصدير النسخة: هنا يتم ترجمة نموذج الأنطولوجية لصيغة يمكن استخدامها والاستفادة منها من قبل خوارزميات التنقيب عن البيانات، أو تصدير قاعدة المعرفة إلى صيغة قاعدة البيانات.

٤. النمذجة: في هذه المرحلة يتم اختيار تقنيات النمذجة المختلفة وتطبيقها واختبارها من أجل التحقق من الأنماط التي تم إيجادها من وجهات نظر مختلفة.

٥. التقييم: تتناول هذه المرحلة التحقق من صحة المعرفة المستخرجة من مرحلة نمذجة التنقيب عن البيانات. ويمكن تجزئة هذه المرحلة لثلاث قنوات:

a. استخراج المعرفة: التركيز في هذه المهمة على تحويل المعرفة المكتشفة من خلال نموذج التنقيب في البيانات لنموذج أنطولوجي.

b. تقييم المعرفة: التركيز هنا على تحليل النموذج الأنطولوجي للتحقق من صحة المعرفة المطلوبة لمزيد من معرفة النطاق للحفاظ على قاعدة المعرفة.

c. تعلم الأنطولوجية: يتم التعامل في هذه المرحلة مع المعرفة المقيمة لتحديث النموذج الأنطولوجي الحالي.

"يهدف الدمج بين هندسة الأنطولوجية واكتشاف المعرفة في البيانات (KDD) إلى نجاح التنقيب في البيانات في المجالات المعقدة والديناميكية" [38]. ويعتبر النموذج (ODKD) واحداً من أوائل النماذج التي تحاول الدمج بين النهجين، قدم Gottgtroy نموذجه في العام ٢٠٠٧ كاقترح للدمج بين هندسة الأنطولوجية واكتشاف المعرفة في البيانات (KDD) [55]، ثم قدم بحثاً متكاملًا حول هذا النموذج كأطروحة قدمت إلى جامعة أوكولاند للتكنولوجيا (AUT)

لنيل درجة الدكتوراه (PhD)، فاعتبر البيانات تمثيلاً للمعرفة الخاصة بنطاق العمل، والمعرفة نقطة انطلاق أي نموذج لاكتشاف المعرفة، "المعرفة المستخرجة من التحليل الجيني لنسيج ميكروي مريض، يمكن الاستفادة منها في إيجاد دواء لمرض آخر مختلف تماماً" [38]. حاول Gottgroy من خلال هذا النموذج الإجابة عن مجموعة أسئلة، أهمها:

- هل يمكن الدمج بين هندسة الأنطولوجية وعمليات اكتشاف المعرفة في البيانات (KDD).
- هل تدعم الأنطولوجية عمليات اكتشاف المعرفة (KD).
- هل تساعد عمليات اكتشاف المعرفة (KD) في بناء الأنطولوجية.

ولكن على الرغم من الإسهامات في مجال اكتشاف المعرفة وتناقيلها المقدمة في هذا النموذج، إلا أنه يفتقر لتوصيف آليات تخزين المعرفة المكتشفة، ونقلها، وإظهارها من جهة، ويفتقر لتوصيف للموارد البشرية المتضمنة في هذا النموذج، وأدوار كل منها من جهة أخرى. سنقوم فيما يلي بتلخيص أبرز سمات هذا النموذج ضمن إطار المقارنة المعرف في هذه الدراسة من خلال الجدول (٨) التالي:

الجدول (٨) سمات النموذج (ODKD, 2010) ضمن إطار المقارنة المعرف

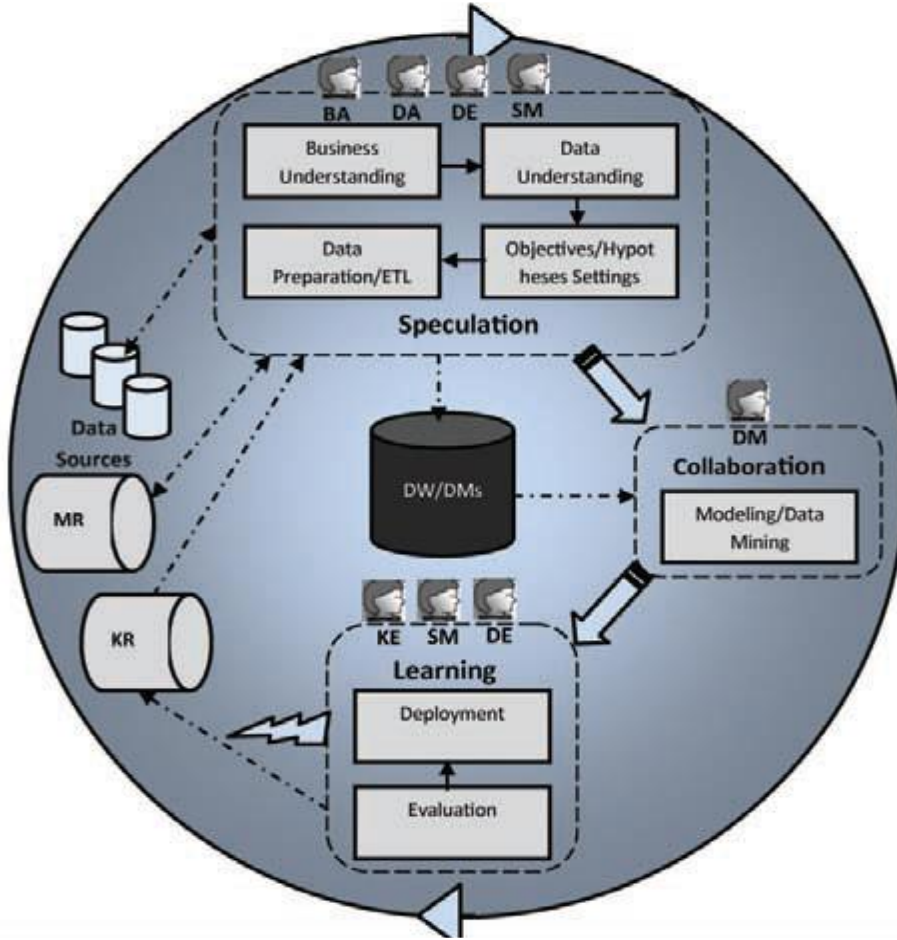
معامل المقارنة	سمات النموذج
النهج المتبع	اتبع نهجاً هجيناً، ويعدّ من أوائل النماذج التي قدمت مفاهيم هندسة الأنطولوجية في نموذج واضح لاكتشاف المعرفة في البيانات.
نوع النموذج والأنظمة التجارية الداعمة	نموذج أكاديمي.
مخازن البيانات	تشكل البيانات الجامدة الدّخل الأساسي للنموذج، دون الاستفادة من مستودع البيانات وأسواقها (DW\DMs). إلا أنه قام بتمثيل هذه البيانات من خلال نماذج الأنطولوجية المختارة.
طبيعة العمليات	تتكوّن دورة حياة النموذج (ODKD, 2010) من خمس مراحل أساسية، هي: (I) تحضير الأنطولوجية. (II) تحليل الأنطولوجية. (III) تحضير النسخة. (IV) النمذجة. (V) التقييم. يتضح من خلال هذه العمليات التي تتضمنها دورة حياة النموذج أنّ Gottgroy اهتمّ بنماذج الأنطولوجية، واختيارها، وتمثيل البيانات من خلالها، وتحليلها، وتحديث هذه النماذج بناءً للمعرفة المكتشفة، دون أن يهتمّ بالبيانات، أو مصادرها. كما لا توجد مرحلة نشر واضحة لنشر النتائج بعد تقييمها والتحقق من صحتها
طبيعة الموارد البشرية	أشار Gottgroy لأهمية الاجتماعات مع العاملين في المجال الذي يتم العمل ضمنه، لتوضيح المتطلبات والأولويات إلى جانب تحديد القيود في بيئة

العمل. ولكنه على الرغم من ذلك لم يشر لأهمية الموارد البشرية في نمودجه، ودور كل منها خلال العمليات التي تضمنها دورة حياة النموذج.	
اتسمت دورة حياة النموذج (ODKD, 2010) بالطابع التسلسلي التتابعي، فلم تتضمن أية حلقات خارجية تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. ولم يذكر أي وجود لمسارات داخلية التي قد تساعد في العودة لعملية سابقة دون الحاجة لتكرار عمليات لم تتغير أي قضايا متعلقة بها.	طبيعة المسارات
يجعل الطابع التسلسلي التتابعي لدورة حياة النموذج موضوع التكيّف مع التغييرات أمراً صعباً. ولم يتم التطرق لأهمية بناء تطبيقات التنقيب في البيانات بشكل متكيف مع التغييرات الخارجية في توثيق النموذج.	التكيف مع التغييرات
يعتبر البعد المعرفي لاكتشاف المعرفة في البيانات (KDD) الهدف الرئيسي الذي دفع Gottgroy للتفكير بدمج مفاهيم هندسة الأنطولوجية مع عمليات اكتشاف المعرفة في البيانات. وعلى الرغم من أنه أشار لتحديث النماذج الأنطولوجية بناءً للمعرفة المكتشفة، إلا أنه لم يشر لآليات تخزين هذه المعرفة، أو نقلها، أو إظهارها، بهدف مشاركتها والاستفادة منها لاحقاً في مشاريع أخرى للتنقيب في البيانات.	البعد المعرفي
أشار Gottgroy لمشاركة العاملين ضمن مجال العمل في تحديد المتطلبات قبل بداية دورة حياة النموذج (ODKD)، ولكنه بالمقابل لم يتطرق أبداً لتحديد أهداف عمليات اكتشاف المعرفة خلالها، أو كيفية الاستفادة من المعرفة المكتشفة في دعم القرارات الاستراتيجية.	البعد الاستراتيجي

٤-٦. "Adaptive Software Development-Business Intelligence (ASD-BI)" (Alnoukari, 2012)

يعدّ هذا النموذج من أوائل النماذج التي قدمت المنهجيات الرشيقة في مجال نمذجة اكتشاف المعرفة في البيانات (KDD)، فقدم Alnoukari في البداية النموذج (ASD-DM, 2008)، ليقوم ببعض التعديلات عليه لاحقاً، ويقدم النموذج (ASD-BI, 2012). وتكمن الفكرة الأساسية وراء هذا التكامل بأنّ المناهج الرشيقة هي الأنسب عندما تكون المتطلبات غير مؤكدة أو متغيرة، وهذا ما يميز معظم تطبيقات استخراج البيانات واكتشاف المعرفة في البيانات (KDD)، التي يتمّ بناؤها وتطويرها في بيئة عمل متغيرة ومتطورة باستمرار حيث من الصعب تطبيق المنهجيات التقليدية.

يتألف هذا النموذج من سبع خطوات أساسية، جُمعت في ثلاث مراحل على شكل دورات متتابعة كما يوضح الشكل (١٩) التالي:



الشكل (١٩) بنية النموذج (ASD-BI, 2012) [7]

سنقوم فيما يلي باستعراض المراحل التي تمر بها دورة حياة النموذج (ASD-BI, 2012)، وتوضيح العمليات

خلالها:

١. التخمين: يتم في هذه المرحلة إجراء كافة مهام بدء المشروع بما في ذلك: تحديد أهداف ومتطلبات كل دورة، وتحديد العدد الأمثل من الدورات وزمن كل دورة، وتحديد البنود ذات المخاطر العالية في وقت مبكر من المشروع. تتألف هذه المرحلة من الخطوات التالية:

a. فهم بيئة العمل: تُشكل نقطة الانطلاق لنموذج ASD-BI، مما يجعل هذه الخطوة واحدة من أهم

عمليات النموذج لأنها تؤثر على كامل المشروع.

- b. فهم البيانات: تُعد إحدى المراحل الحيوية في أي مشروع من مشاريع ذكاء الأعمال، ففي هذه الخطوة تصبح البيانات مألوفة للمستخدمين مما يساعد في معالجة كافة القضايا المتعلقة بالبيانات في المراحل اللاحقة.
- c. إعداد الفرضيات والأهداف: يتم في هذه الخطوة استخدام نتائج المرحلتين السابقتين لتحديد أهداف تطبيق ذكاء الأعمال (BI) ووضع الفرضيات.
- d. إعداد البيانات / ETL: تُعد أهم خطوة في هذه المرحلة فهي تتضمن جميع العمليات اللازمة لبناء عينة البيانات التي سيتم استخدامها في المراحل اللاحقة وذلك بالاستفادة من الخطوات السابقة.
٢. التعاون: الاهتمام في هذه المرحلة بتحقيق التعاون والتواصل بشكل فعال بين أفراد فريق العمل، بحيث يتم تبادل الخبرات والمعارف التي تمكنهم من إدارة المشروع واتخاذ القرارات وحل المشاكل في المستقبل. تتكون هذه المرحلة من خطوة واحدة هي:
- a. النمذجة / التنقيب في البيانات: تهدف هذه الخطوة إلى اعتماد نموذج فعال للعمل أو اختيار إحدى خوارزميات التنقيب في البيانات، بالاعتماد على أهداف المشروع وطرق التنقيب في البيانات: التصنيف (Classification) والعنقدة (Clustering) والتجميع (Association) والتسلسلية (Sequencing) والتنبؤ والتوقع (Forecasting).
٣. التعلم: يتم التركيز في هذه المرحلة على استعراض الجودة من وجهات نظر مختلفة: وجهة نظر العملاء من جهة والتقنيات المستخدمة وحالة المشروع من جهة أخرى. فوجهة نظر العملاء ورؤيتهم هي المحور الأساسي في مرحلة التعلم، فيما تشكل المراجعة من الناحية التقنية عامل هام من أجل الحصول على منتج عالي الجودة. تتألف هذه المرحلة من الخطوتين التاليتين:
- a. التقييم: تُعد هذه الخطوة إحدى الخطوات الحيوية التي تضمن أن طريقة النمذجة أو خوارزمية التنقيب عن البيانات المختارة مناسبة لتحقيق أهداف المشروع. ليتم تخزين النتائج الصحيحة التي تم الحصول عليها من عملية النمذجة وخوارزميات التنقيب في البيانات في مخازن المعرفة (KR).
- b. النشر: تمثل نهاية كل دورة من دورات ASD-BI، حيث يجب تنظيم المعرفة المكتشفة وعرضها بالشكل الذي يمكن العملاء من استخدامها، فيمكن عرض هذه المعرفة المكتشفة من خلال وسائل متعددة: لوحات القيادة والتقارير التحليلية والتنبهات وغيرها. ليتم بعدها تقييم المشروع ونتائج الدورة الحالية، والتّحضير للدورة القادمة.

يعدّ النموذج (ASD-DM, 2008) - كما أسلفنا سابقاً- واحداً من أوائل المحاولات للدمج بين المنهجيات الرشيقة ونمذجة اكتشاف المعرفة في البيانات (KDD)، إلا أنّ هذا النموذج كان يواجه بعض القيود والقصور في بعض الجوانب، من أهمها: لا تحديد لمصادر البيانات بشكل واضح، وعدم وجود مرحلة نشر واضحة، وتجاهل إشراك أي نوع من أنواع الموارد البشرية، بالإضافة لعدم تخزين المعرفة المكتشفة في أيّ مكان. ليتمّ البناء عليه في هذا النموذج (ASD- BI, 2012) وتلافي نقاط القصور تلك. يسعى Alnoukari من وراء اعتماد المنهجيات الرشيقة في بناء تطبيقات ذكاء الأعمال (BI) للتقليل من تكاليف عمليات التعديل والتكيف مع المتغيرات في المتطلبات وبيئات العمل، "فهذه المنهجيات ملائمة جداً في ظروف بيئات العمل المتغيرة والمتسارعة، حيث من الصعب استخدام المنهجيات التقليدية في مثل هذه الأسواق والبيئات المتطورة وغير المستقرة" [7].

في الواقع لا يزال هذا النموذج يواجه بعض نقاط الضعف، بشكل أساسي في طريقة بناء مخازن المعرفة (KR)، وطريقة تنظيم المعرفة المكتشفة، ومشاركتها، أو آلية استخدام هذه المعرفة لمتابعة استراتيجيات العمل وتعديلها، كما أنه يقوم على مشاركة ضخمة للموارد البشرية والخبراء وكأنّه أعدّ للمشاريع الكبيرة من مشاريع ذكاء الأعمال (BI) والتنقيب في البيانات فقط.

سنحاول تلخيص أبرز سمات هذا النموذج ضمن إطار المقارنة المعرف في هذه الدراسة من خلال الجدول (٩)

التالي:

الجدول (٩) سمات النموذج (ASD-BI, 2012) ضمن إطار المقارنة المعرف

معامل المقارنة	سمات النموذج
النهج المتبع	اتبع نهجاً نهجياً يقوم على الدمج بين المنهجيات الرشيقة وعمليات اكتشاف المعرفة في البيانات (KDD).
نوع النموذج والأنظمة التجارية الداعمة	نموذج أكاديمي.
مخازن البيانات	تتكوّن مخازن البيانات ومصادرها في النموذج (ASD-BI, 2012) من: (I) البيانات المجردة. (II) مستودع البيانات وأسواقها (DW\DMS). (III) مخزن البيانات المترفعة (Meat data Repository). (IV) مخزن المعرفة (KR). تشكل البيانات المجردة ومستودع البيانات وأسواقها (DW\DMS) الدّخل الأساسي للنموذج، حيث تشكل البيانات المجردة دخل المرحلة الأولى (التّخمين)، ويتم بناء مستودع البيانات وأسواقها (DW\DMS) وتحويل البيانات من مصادرها إليها خلال مرحلة التّخمين، ليشكل هذا المستودع دخل المرحلة الثانية من دورة حياة التّموذج. ويتمّ تخزين

<p>البيانات المترفعة (Metadata) الناتجة عن عملية فهم البيانات في هذا مخزن البيانات المترفعة (Metadata Repository). فيما يشكل مخزن المعرفة (KR) الخرج الأساسي للنموذج، ويتم تخزين المعارف المكتشفة ضمنه للاستفادة منها في دورات النموذج اللاحقة أو في مشاريع أخرى.</p>	
<p>تتكوّن دورة حياة النموذج (ASD-BI, 2012) من سبع عمليات أساسية وهي: (I) فهم بيئة العمل. (II) فهم البيانات. (III) وضع الأهداف وتحديد الفرضيات. (IV) تحضير البيانات/ETL. (V) النمذجة والتنقيب في البيانات. (VI) التقييم. (VII) النشر. استفاد Alnoukari من النموذجين (CRISP-DM, 2008) و (DMLC, 2009) في وضع بنية عمليات تفصيلية وشاملة خلال دورة حياة النموذج (ASD-BI)، ولكن من الملاحظ أنّ عملية تحضير البيانات تتطلب جهداً كبيراً بالمقارنة مع العمليات الأخرى، بما تتضمنه من مهام ووظائف اختيار البيانات المنشودة، وتنظيفها، ومعالجتها، وتحويلها لمستودع البيانات وأسواقها (DW\DMs).</p>	<p>طبيعة العمليات</p>
<p>يتضمن النموذج (ASD-BI, 2012) الموارد البشرية التالية: (I) محلّل الأعمال. (II) محلل البيانات ومهندسها. (III) المدير الاستراتيجي. (IV) المنقب في البيانات. (V) مهندس المعرفة. (VI) خبير مجال العمل. يتميز النموذج بتوصيفه الجيد للأفراد والخبرات المضمنة في بنيته، وتوصيف مهمة كل منها ودورها خلال دورة حياته. ولكن على الرغم من إشارته لأنّ الموارد البشرية السابقة تشير لوظائف ومهارات كل منها وليس أفراد بحد ذاتهم، إلاّ أنّه قدم محلل البيانات ومهندس البيانات في دور واحد هو محلل البيانات ومهندسها (Data analyst\ engineer). فمحلل البيانات يعتبر مسؤولاً عن الوصول لمصادر البيانات، وتحليل محتواها وبنيتها، ويقوم بتوصيفها، وتحديد درجة جودتها. بالمقابل يملك مهندس البيانات خبرة كافية في مجال تصميم وبناء مستودع البيانات وأسواقها (DW/DMs)، تجعله مسؤولاً عن عمليات اختيار عينات البيانات، ومعالجتها، وتحويلها، مستودع البيانات وأسواقها (DW/DMs).</p>	<p>طبيعة الموارد البشرية</p>
<p>تمّ دورة حياة النموذج (ASD-BI, 2012) بعدة دورات تسمح بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة. بالمقابل لم يتم الحديث عن الحاجة لوجود مسارات داخلية حتى مرحلة التقييم، بحيث يتم</p>	<p>طبيعة المسارات</p>

<p>العودة بدورة أخرى لتصحيح النتائج، على الرغم من عدم حدوث أية تغييرات على مستوى عمليات أخرى كفهم بيئة العمل أو فهم البيانات.</p>	
<p>تُعتبر قضية التكيف مع التغييرات خلال دورة حياة النموذج (ASD-BI, 2012) الهدف الأساسي من وراء تقديم المنهجية الرشيقة (ASD: Adaptive Software Development) ضمن دورة حياة النموذج، بما "تتصف به هذه المنهجيات من مجموعة مبادئ مثل: التركيز على العميل، والتخطيط، والإصدارات الصغيرة، الاستعارة، الترميز المعياري، التكامل المستمر، وألعاب التخطيط" (Alnoukari, 2012).</p>	<p>التكيف مع التغييرات</p>
<p>اهتمّ النموذج (ASD-BI, 2012) بعمليات تحصيل المعرفة، وإظهارها. فيتم تخزين المعرفة المكتشفة في مخزن المعرفة (KR). ويسهم مهندس المعرفة المضمن في هذا النموذج في عمليات تحصيل المعارف والمعلومات المكتشفة، وهيكلتها بصيغة معيارية لتخزينها وإظهارها. ولكن مازال يواجه بعض القصور في توضيح آليات تخزين المعرفة المكتشفة وبناء مخزن المعرفة (KR).</p>	<p>البعد المعرفي</p>
<p>أشار Alnoukari لأهمية تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات في كشف فرص تنافسية في سوق العمل، فاحتوت بنية النموذج على مرحلة واضحة لوضع أهداف هذه التطبيقات، وتحديد الفرضيات والقيود المتعلقة ببيئة العمل. يلعب المدير الاستراتيجي دوراً أساسياً خلال هذه المرحلة في صياغة استراتيجيات المؤسسة.</p>	<p>البعد الاستراتيجي</p>

٥. مقارنة بين النماذج المختارة

٥-١. ملخص نقاط الضعف والقوة للنماذج المختارة

يلخص الجدول (١٠) أهم نقاط القوة والضعف في النماذج الستة التي تم مناقشتها سابقاً في هذا الفصل، وذلك بناءً على معاملات إطار المقارنة المعرفة سابقاً، هذه المعاملات هي: (I) النهج المتبع. (II) نوع النموذج والأنظمة التجارية الداعمة. (III) تحديد طبيعة مصادر البيانات ومخازنها. (IV) تضمين الموارد البشرية وطبيعتها. (V) طبيعة العمليات. (VI) المسارات التي تتضمنها دورة حياة النموذج. (VII) التكيف مع التغييرات ومراعاتها لبيئات العمل المتطورة باستمرار. (IX) البعد المعرفي ودعمها لعمليات تحصيل المعرفة، وتخزينها، ومشاركتها. (X) البعد الاستراتيجي.

معظم النماذج التي تم اختيارها تعتمد الخطوات الأساسية التالية نفسها: (I) تحديد مجال العمل. (II) تحضير البيانات. (III) التنقيب في البيانات. (IV) التقييم (V) النشر.

الجدول (١٠) ملخص نقاط القوة والضعف في نماذج اكتشاف المعرفة المختارة في دراستنا

نقاط الضعف	نقاط القوة	النهج المُتبَع	نموذج اكتشاف المعرفة في البيانات (KDD)
<ul style="list-style-type: none"> • البيانات هو المصدر الوحيد، لا يوجد استخدام صريح لمستودع البيانات أو أسواقها (DW\DMs). "إنّ استخدام مستودع البيانات وأسواق البيانات يوفر قاعدة صلبة من البيانات لاستخدامها في مرحلة التنقيب في البيانات، ويوفر سرعة التنفيذ لمشاريح التنقيب في البيانات" [56]. • لا وجود لمرحلة "النشر" مما يجعل من الصعب تقييم واختبار نتائج مرحلة التنقيب في البيانات. • لا وجود لدورة داخلية حتى نهاية مرحلة التقييم. • الافتقار للاكتمال والسلامة بما يضمن النجاح لمشروع التنقيب في البيانات. • الافتقار لمنظور بيئة العمل. • تجاهل إشراك الموارد البشرية. • صعوبة التكيف مع متطلبات التغييرات. • لا يتم تخزين المعرفة المكتشفة في أي مكان. 	<ul style="list-style-type: none"> • يعدّ حجر الزاوية لجميع نماذج اكتشاف المعرفة (KDD) اللاحقة. • مراحل عمل واضحة وبسيطة. • عزف المراحل الأساسية الأولى لاكتشاف المعرفة في البيانات (KDD) (KDD). • مراحل اختيار البيانات ومعالجتها وتحويلها مفصلة بشكل جيّد. 	النهج التقليديّ	Fayyad et al (١٩٩٦)
<ul style="list-style-type: none"> • نقاط ضعف نموذج فياض ذاتها. 	<ul style="list-style-type: none"> • يمكن الاعتماد عليه باعتباره نموذجاً يقوم على الممارسات التحليلية الحديثة. • اعتماده على عينات من البيانات يجعل منه فعالاً من حيث الوقت والتكلفة. 	النهج التقليديّ	SEMMA (١٩٩٧)
<ul style="list-style-type: none"> • البيانات هو المصدر الوحيد، لا يوجد استخدام صريح لمستودع البيانات أو أسواقها (DW\DMs). • تعدّ مرحلة النشر نقطة نهاية المشروع، فينتهي المشروع بمجرد إعداد التقرير النهائي. • تجاهل الموارد والخبرات البشرية. • المعرفة المكتشفة لا يتم تخزينها في أيّ مكان. 	<ul style="list-style-type: none"> • يعدّ واحداً من أكثر نماذج اكتشاف المعرفة في البيانات (KDD) استخداماً في مجال التنقيب عن البيانات. • مراحل معرفة بشكل جيد تجعل منه سهل الاستخدام. • تجزئة المراحل الأساسية لمراحل تفصيليّة ضمنيّة. • توثيق جيد. 	النهج التقليديّ	CRISP-DM (٢٠٠٨)
<ul style="list-style-type: none"> • صعب التحقيق بما يحتاجه من موارد هائلة. • عملية تخزين المعرفة المكتشفة غير واضحة. • توثيق ضعيف. 	<ul style="list-style-type: none"> • استخدام مصادر بيانات متعددة: مستودع البيانات وأسواقها (DW\DMs) البيانات، ومخازن المعرفة والمعلومات (IKR). • مراحل معرفة بشكل جيد تجعل منه سهل الاستخدام. • تجزئة المراحل الأساسية لمراحل تفصيليّة ضمنيّة. 	النهج التقليديّ	Hoffmann (٢٠٠٩)

	<ul style="list-style-type: none"> تضمنين جيّد للموارد والخبرات البشرية. المعرفة المكتشفة يتم تخزينها في مخازن المعرفة والمعلومات (IKR). 		
<ul style="list-style-type: none"> آلية تدفق البيانات غير منتظمة. عدم وضوح عملية تخزين المعرفة المكتشفة. تضمنين ضعيف للموارد والخبرات البشرية. 	<ul style="list-style-type: none"> يعتبر واحداً من أوائل النماذج التي حاولت الدمج بين مفاهيم هندسة الأنطولوجية وعمليات اكتشاف المعرفة في البيانات. التحقق والتثبت من المعرفة المكتشفة. دعم الوصف الدلالي والمخطط. دعم تحديثات المعرفة النشطة. 	النّهج الذي يعتمد على هندسة الأنطولوجيّة	Gottgory (2010)
<ul style="list-style-type: none"> لا وجود لدورات داخلية حتى نهاية مرحلة التّشّ. صعب التحقيق بما يحتاجه من موارد هائلة. عملية تخزين المعرفة المكتشفة غير واضحة، فما يزال يواجه قصوراً في طريقة بناء مخازن المعرفة من جهة وتخزين المعرفة المكتشفة وتنظيمها من جهة أخرى. 	<ul style="list-style-type: none"> يعتبر التقري من أوائل من حاول إدخال المنهجيات الرّشيقية في مجال نمذجة اكتشاف المعرفة في البيانات (KDD). مراحل تكرارية بسيطة وواضحة. الاستجابة بسهولة للتغيرات في متطلبات بيئة العمل والأسواق المتطورة بشكل متسارع. تضمنين جيّد للموارد والخبرات البشرية. تحضير البيانات بشكل جيد، والاعتماد على مستودع البيانات وأسواقها (DW\DMs). 	النّهج الذي يعتمد على المنهجيات الرّشيقية	التّقري (2012)

٥-٢. مصفوفة السمات والخصائص المميزة

يبين الجدول (١١) مصفوفة السمات والخصائص المميزة للنماذج التي تمّ اختيارها في دراستنا هذه، وتموضع حلنا المقترح في هذا البحث بالنسبة لها، بحيث يظهر الاختلافات بين نماذج اكتشاف المعرفة في البيانات (KDD) استناداً للمعايير التي بُني عليها إطار المقارنة: (I) النهج المتبع. (II) الأنظمة التجارية الدّاعمة. (III) مخازن البيانات. (IV) طبيعة العمليات. (V) طبيعة الموارد البشريّة. (VI) طبيعة المسارات. (VII) التكيّف مع التغيرات. (VIII) البعد المعرفي. (IX) البعد الاستراتيجي.

الجدول (١١) مصفوفة السمات والخصائص المميزة لنماذج اكتشاف المعرفة التي تمّ اختيارها في دراستنا

SCRUM-BI (2015)	ASD-BI (2012)	ODKD (2007)	DMLC (2009)	CRISP-DM (2008)	SEMMA (1997)	KDD (1996)	
							عام
أكاديمي	أكاديمي	أكاديمي	أكاديمي	صناعي	صناعي	أكاديمي	نوع النموذج
لا يوجد	لا يوجد	لا يوجد	لا يوجد	IBM SPSS Modeler	SAS Enterprise Miner™	MineSet™	الأنظمة التجاريّة الدّاعمة
هجين	هجين	هجين	تقليدي	تقليدي	تقليدي	تقليدي	النّهج المتبع

المنهجيات الرشيقة	المنهجيات الرشيقة	مفاهيم الأنطولوجية	-	-	-	-	المفاهيم أو الأدوات المستخدمة
SCRUM	ASD	-	-	-	-	-	المنهجية المستخدمة
العمليات الدّاخلية	الدورات الخارجية	-	-	-	-	-	مستوى تطبيق المنهجية
							مخازن البيانات
√	√	√	√	√		√	الاهتمام بمخازن البيانات
مستودع البيانات وأسواقها	البيانات المجردة ومستودع البيانات وأسواقها	البيانات المجردة	مستودع البيانات وأسواقها	البيانات المجردة	غير محدد	البيانات المجردة	مصادر البيانات
							طبيعة العمليات
٧	٧	٥	٩	٦	٥	٥	عدد العمليات
√	√		√	√			فهم بيئة العمل
√	√		√	√			فهم البيانات
		√					تحضير الأنطولوجية
		√					تحليل الأنطولوجية
		√					تحضير النسخة
√	√		√				وضع الأهداف والفرضيات
√			√		√	√	اختيار البيانات/أخذ العينات
			√		√	√	معالجة البيانات
	√			√			تحضير البيانات
			√		√	√	تحويل البيانات
√	√	√	√	√	√	√	النمذجة/التنقيب في البيانات
√	√	√	√	√	√	√	التقييم والاختبار
√	√		√	√			النشر
							طبيعة الموارد البشرية
√	√		√				الاهتمام بدور الأفراد
٨	٦		٨	لا يوجد	لا يوجد	لا يوجد	عدد الأدوار المشاركة بالنموذج
			√				مدير المشروع
√							خبير المنهجية
√	√		√				محلل الأعمال

√			√				محلل البيانات
√	√		√				مهندس البيانات
√	√		√				المنقب بالبيانات
√	√		√				خبير مجال العمل
√	√		√				مدير الاستراتيجية
√	√		√				مهندس المعرفة
							البعد المعرفي
√	√	√	√				الاهتمام بالمعرفة
√	√		√				مخازن المعرفة والمعلومات
							طبيعة المسارات
√	√		√	√	√		دورات خارجية
√			√	√		√	مسارات داخلية
√	√	√	√	√	√	√	تدفق تسلسلي تنبؤي
√			√				تدفق ثنائي الاتجاه
							التكيف مع التغيرات
√	√		√	√			الاهتمام بالتكيف مع المتغيرات
							البعد الاستراتيجي
√	√		√				الاهتمام بالاستراتيجية

٦. تموضع الحل المقترح

تبيّن الدراسة التحليلية لنماذج اكتشاف المعرفة في البيانات (KDD) التي تمّ اختيارها في هذا البحث بعض نقاط الضعف والقصور فيها، والتي تم تلخيصها من خلال الجدول (١١) جدول السمات والخصائص المميزة. بالاستفادة من الدراسة السابقة يمكن وضع قائمة بالمبادئ والسمات الأساسية للحل المقترح، والتي توضح المساهمات المرجوة في هذا البحث، وتموضعه بالنسبة للدراسات السابقة:

١. يتبع الحلّ المقترح نهجاً هجيناً لاكتشاف المعرفة في البيانات (KDD)، يقوم على الدمج والمزاوجة بين مفاهيم اكتشاف المعرفة في البيانات (KDD) التقليدية وخصائص ومميزات المنهجية الرشيقية سكروم في التكيف مع المتغيرات على مستوى المتطلبات وبيئات العمل -والذي يعتبر همّنا الأساسي في الحلّ المقدم- والاستفادة مما تقدمه هذه المنهجية من أدوات وقدرات لإدارة الأحداث والأفراد المشاركين بالعمل.

٢. تمرّ دورة حياة الحلّ المقترح بعدد من الدورات، يتمّ تحديدها في بداية المشروع، بهدف بناءه على مراحل تمكن من التسليم السريع من جهة، والاستفادة من الخبرات المتراكمة خلال الدورات السابقة على مستوى كل من البيانات وبيئة العمل من جهة أخرى. بالإضافة لاحتوائه على مسارات داخلية، تسمح بالاستفادة من ردود أفعال المشاركين بالعمل، والتكيف معها، وتصحيح عمليات التحليل والتنقيب في المعطيات دون العودة لعمليات أخرى لن يحصل فيها أيّ تغيير، أو انتظار الدورات اللاحقة للمشروع كما في بعض النماذج السابقة التي لم تتضمن مثل هذه الخيارات، ومنها النموذجين: (ASD-BI, 2012; ODKD, 2010).

٣. يعتمد الحلّ المقترح على مستودع البيانات وأسواقها (DW\DMs) كمصدر البيانات الوحيد، والتي "تعتبر مصدراً للبيانات لا مفر منه في تطبيقات التنقيب في البيانات، والتي تضمن جودة المعلومات والمعارف المكتشفة" [10]، فمعظم النماذج السابقة لم تكن واضحة في تعريف مصدر البيانات: فالبعض اعتمد على أي مصدر للبيانات، كالنموذجين: (ODKD, 2010; ; KDD, 1996)، والبعض الآخر لم يتطرق نهائياً لهذا الموضوع خلال دورة حياة النموذج كالنموذج (CRISP-DM, 2008)، فيما اعتبر آخرون البيانات المجردة إلى جانب مخازن المعطيات مصدراً للبيانات كالنموذج (ASD-BI, 2012)، ويتم خلال مراحل تحضير البيانات من دورة حياة النموذج تحويل البيانات لمخازن المعطيات من مصادرها المختلفة، وما يتضمنه ذلك من عمل شاق وطويل يتم القيام به خلال إحدى العمليات في دورة حياة النموذج.

٤. يهتمّ الحلّ المقترح بالمعرفة المكتشفة وآليات تخزينها ونقلها وإظهارها، بالاعتماد على مخازن المعرفة والمعلومات (IKR) لتخزين المعارف والمعلومات المستخرجة، وتناقلها بين الأفراد خلال دورة حياة الحلّ المقترح. إنّ النموذجين (DMLC, 2009; ASD-BI, 2012) هما الوحيدين الذين أشارا لاستخدام مخازن المعرفة والمعلومات، وأهمية تخزين المعارف والمعلومات الجديدة.

٥. تتكون دورة حياة الحلّ المقترح من ٧ عمليات أساسية: فهم البيانات، فهم بيئة العمل، وضع الأهداف والفرضيات، اختيار البيانات وتحويلها، النمذجة والتنقيب في البيانات، التقييم، النشر. تمّ اختيار هذه العمليات بحيث تكون واضحة التسمية والمعنى، وتفصيلية وشاملة لكافة العمليات اللازمة لتحليل البيانات نمذجتها، وعمليات التحضير وفهم البيانات ومجال العمل السابقة لها، وعمليات التقييم والاختبار والنشر اللاحقة لها من جهة، ولا تسبب ضياعاً أو تشويشاً في إدارة دورة حياة النموذج. فبعض النماذج السابقة احتوت على عدد قليل من العمليات، كالنموذجين (KDD, 1996; SEMMA, 1997) وأغفلت مراحل مهمة في اكتشاف المعرفة في البيانات (KDD) والمتعلقة بفهم كل من البيانات ومجال العمل، بالمقابل احتوت نماذج أخرى على عدد كبير من العمليات، كالنموذج (DMLC, 2009) حتى أنّ بعض العمليات هي أقرب لمهمة أو نشاط

صغير منه إلى عملية مجد ذاتها، كعملية تحويل البيانات التي تقوم على تطبيق عمليات التقليل والإسقاط (data reduction and projection) لا أكثر.

٦. تتكون الموارد البشرية المشاركة في الحلّ المقترح من تسع أفراد، وهم: (I) المدير الاستراتيجي. (II) خبير المنهجية. (III) محلل الأعمال. (IV) محلل البيانات. (V) مهندس البيانات. (VI) المنقب في البيانات. (VII) خبير مجال العمل. (VIII) مهندس المعرفة. إنّ النموذجين (DMLC, 2009; ASD-BI, 2012) هما الوحيدين الذين أشارا للأفراد المشاركة في مراحل النموذج ودور كل منها، فتلعب الموارد البشرية دوراً أساسياً في أي عمل يقوم على اكتشاف المعرفة، وتبادلها، وإدارتها.
٧. يركز الحلّ المقترح على البعد الاستراتيجي لتطبيقات ذكاء الأعمال (BI) والتنقيب في المعطيات، من خلال تضمين المدير الاستراتيجي في عملية وضع الأهداف والفرضيات لهذه التطبيقات، بالشكل الذي يساعد في صياغة الخطط الاستراتيجية وتحقيقها.

٧. خاتمة

قدمنا في هذا الفصل مقارنة بين أهم نماذج اكتشاف المعرفة في البيانات (KDD) استناداً للمعاملات ضمن إطار المقارنة الذي عرفناه سابقاً في هذا الفصل، كما قمنا بتصنيف هذه النماذج في فئتين أساسيتين بحسب النهج المتبع، وهما: النهج التقليدي، والنهج الهجين، ومن الواضح أنّ معظم نماذج اكتشاف المعرفة في البيانات (KDD) تتبع النهج التقليدي. ثم قمنا ببناء مصفوفة السمات والخصائص المميزة للنماذج المختارة، مما يسمح بالاستفادة من نقاط القوة فيها والعمل على تلافي نقاط الضعف في الحلّ المقترح في هذا العمل.

الفصل الرابع

النموذج (SCRUM-BI) لاكتشاف المعرفة في البيانات

١. مقدمة

يواجه صانعو القرار صعوبات كبيرة في معالجة البيانات الهائلة التي تنشأ عن الأنظمة التشغيلية (CRM، ERP، EIS، ...)، مما يمنع الوصول إليها والاستفادة منها بشكل فعال، ما أدى لظهور نظم دعم القرار (DSS) وتطبيقات ذكاء الأعمال (BI) لاكتشاف المعرفة في البيانات، من خلال ما تؤمنه من أدوات لتنظيم عمليات تخزين البيانات، ونقلها، والوصول إليها، ومعالجتها، وتحليلها، والتنقيب فيها، بهدف الحصول على معارف ومعلومات مفيدة تساعد في اتخاذ القرار المناسب في الوقت المناسب. كل ذلك أدى لزيادة الاهتمام بتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات بشكل ملحوظ خلال السنوات الأخيرة، ولكن هذه التطبيقات لازالت تواجه العديد من التعقيدات والصعوبات بسبب التطور السريع لبيئات العمل والأسواق، والتغيرات الديناميكية المستمرة على مستوى المتطلبات والأولويات، ووجود العديد من أصحاب المصلحة المشاركين في العمل، بالإضافة للجداول الزمنية غير الواضحة، وتعدد الأساليب والنظم لاستخراج البيانات من مصادرها، والمشاكل التي تعترض توافر المعلومات وجودتها.

يشابه الواقع الحالي لتطبيقات ذكاء الأعمال (BI) ونظم دعم القرار (DSS) ومعدلات فشلها المرتفعة التي تتجاوز ٥٠% أزمة صناعة البرمجيات وتطويرها أواخر القرن الماضي، وما واجهته من مشاكل وصعوبات. قدمت المنهجيات الجديدة لتطوير البرمجيات (المعروفة بالمنهجيات الرشيقية) حلولاً فعالة لمواجهة تلك المشاكل والصعوبات، والذي شكل حافزاً للاستفادة من خصائص هذه المنهجيات في التكيف والتعامل مع المتغيرات على مستوى المتطلبات وبيئات العمل.

نقدم في هذا البحث نموذجاً جديداً لاكتشاف المعرفة في البيانات (KDD)، قمنا بتسميته "SCRUM-BI". نراعي في النموذج المقترح الاستفادة من نقاط قوة النماذج السابقة، وتلافي نقاط الضعف فيها، والتي تم توضيحها

ومناقشتها في فصل سابق من هذه الدراسة. يقوم النموذج (SCRUM-BI) على الدمج والمزاوجة بين النهج التقليدي لاكتشاف المعرفة في البيانات (KDD) والمفاهيم والأدوات التي توفرها المنهجية الرشيقية سكروم (SCRUM)، للاستفادة من خصائص هذه المنهجية في التكيف والتجاوب مع المتغيرات الخارجية في المتطلبات وبيئات العمل، وتنظيم الخبرات والمعارف المتراكمة وتناقلها بين أفراد فريق العمل.

سنقوم في هذا الفصل بتوضيح كيفية ملائمة المنهجية سكروم في النموذج SCRUM-BI، موضحين أهمية اعتماد المنهجية سكروم في بناء تطبيقات ذكاء الأعمال (BI) ناجحة. ومن ثم سنفصل بنية النموذج (SCRUM-BI)، وتفصيل عناصره المتمثلة بالعمليات والأفراد ومخازن البيانات، لنستعرض أخيراً خصائص هذا النموذج، ودوره في المساعدة ببناء تطبيقات ذكاء الأعمال (BI) بشكل متكيف ومتجاوب مع المتغيرات في المتطلبات وبيئات العمل، وتوجيه أهداف هذه التطبيقات نحو تحقيق الخطط الاستراتيجية، وتنظيم عمليات الحصول على المعرفة، وتخزينها، ونقلها، وإظهارها.

٢. ملائمة المنهجية الرشيقية سكروم مع اكتشاف المعرفة في البيانات

بعد أن قمنا بتحليل هندسية المنهجية سكروم ومقارنتها بالمنهجيات الرشيقية الأخرى لتحديد سبب اختيارنا لهذه المنهجية، يبقى السؤال الأهم هو: كيفية ملائمة المنهجية في النموذج المقترح في دراستنا هذه، وما يتضمنه من عمليات خلال دورة حياة هذا النموذج.

أحد الأفكار في هذا المجال اعتبار أي تقرير من التقارير النهائية التي يتضمنها التطبيق كأصغر جزء من العمل، بحيث يتم التخطيط وإنجاز كافة الأعمال الخاصة بهذا التقرير في شوط واحد، وفي الشوط الثاني يتم العمل على تقرير آخر، وهكذا حتى يتم إنجاز كافة التقارير التي يتضمنها التطبيق الذي نقوم ببنائه. ولكن في الحقيقة أنّ هذه الطريقة لن تعمل على النحو الأمثل، فمراحل بناء التقرير مهما صغر حجمه وحجم العمل اللازم لبنائه لا يمكن تضمينها في شوط واحد يمتدّ لأربع أسابيع على الأكثر، والأهم من ذلك إن كان التقرير الذي قمنا ببنائه لا يناسب متطلبات الزبون أو المستخدم النهائي مما يتطلب إعادة العمل بكامله بما يتضمنه من تنظيم للبيانات، وبناء مستودع البيانات وأسواقها (DW\DMS)، ونمذجة هذه البيانات، وتطبيق خوارزميات معينة للتنقيب في البيانات، وتصميم التقرير وبنائه. كل ذلك يجعلنا بحاجة لمراجعة العمل والحصول على ردود أفعال الزبائن والمستخدمين النهائيين في وقت سابق، قبل القيام بالكثير من العمل.

لذلك قمنا باعتماد الطريقة الأخرى في النموذج المقترح في هذه الدراسة، وهي ملائمة الأشواط على مستوى كل عملية من العمليات التي يتضمنها النموذج SCRUM-BI خلال دورة حياته كما يوضح الشكل (٢٠)، بحيث تصبح كل دورة من دورات النموذج SCRUM-BI مكونة من أشواط السبع التالية: (I) شوط فهم البيانات. (II) شوط فهم

بيئة العمل. (III) شوط وضع الأهداف والفرضيات. (IV) شوط اختيار عينات البيانات. (V) شوط النمذجة والتنقيب في البيانات. (VI) شوط الاختبار والتقييم. (VII) شوط النشر.

ومن الواضح أنّ هذه الطريقة أسهل وأفضل لملائمة العمل الذي يتم خلال كل عملية ضمن شوط يمتدّ بين أسبوع وأربعة أسابيع، مع إمكانية بناء نماذج للتقارير المتوقعة في نهاية تطبيقات ذكاء الأعمال (BI) خلال شوط وضع الأهداف والفرضيات، بحيث يمكن مراجعتها مع المستخدم النهائي أو الزبون في نهاية هذا الشوط، والحصول على ردود الأفعال في وقت مبكر من مراحل التطوير، وقبل الانتهاء من بناء التطبيق وما يتضمنه من تقارير، بحيث يصبح جميع الأفراد المشتركين بالعمل وأصحاب المصلحة على علم ودراية كافية عن الخرج المتوقع في نهاية كل دورة من دورات النموذج -SCRUM-BI في هذه الدراسة. وبمجرد الموافقة على النماذج التي تمّ مراجعتها يمكن الانتقال لاستكمال عمليات تطوير تطبيق ذكاء الأعمال.



الشكل (٢٠) بنية الأشواط خلال دورة حياة النموذج "SCRUM-BI"

٣. بنية النموذج SCRUM-BI

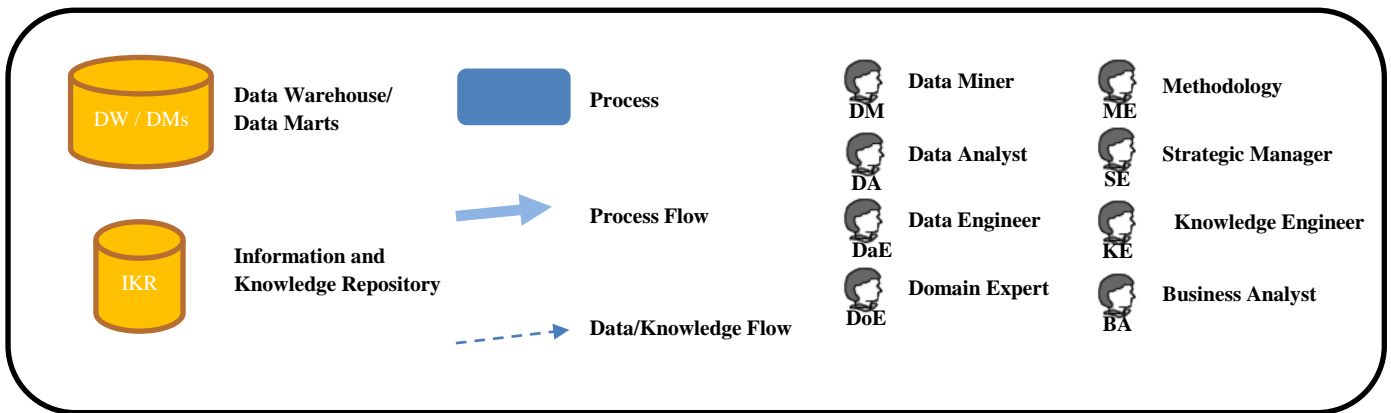
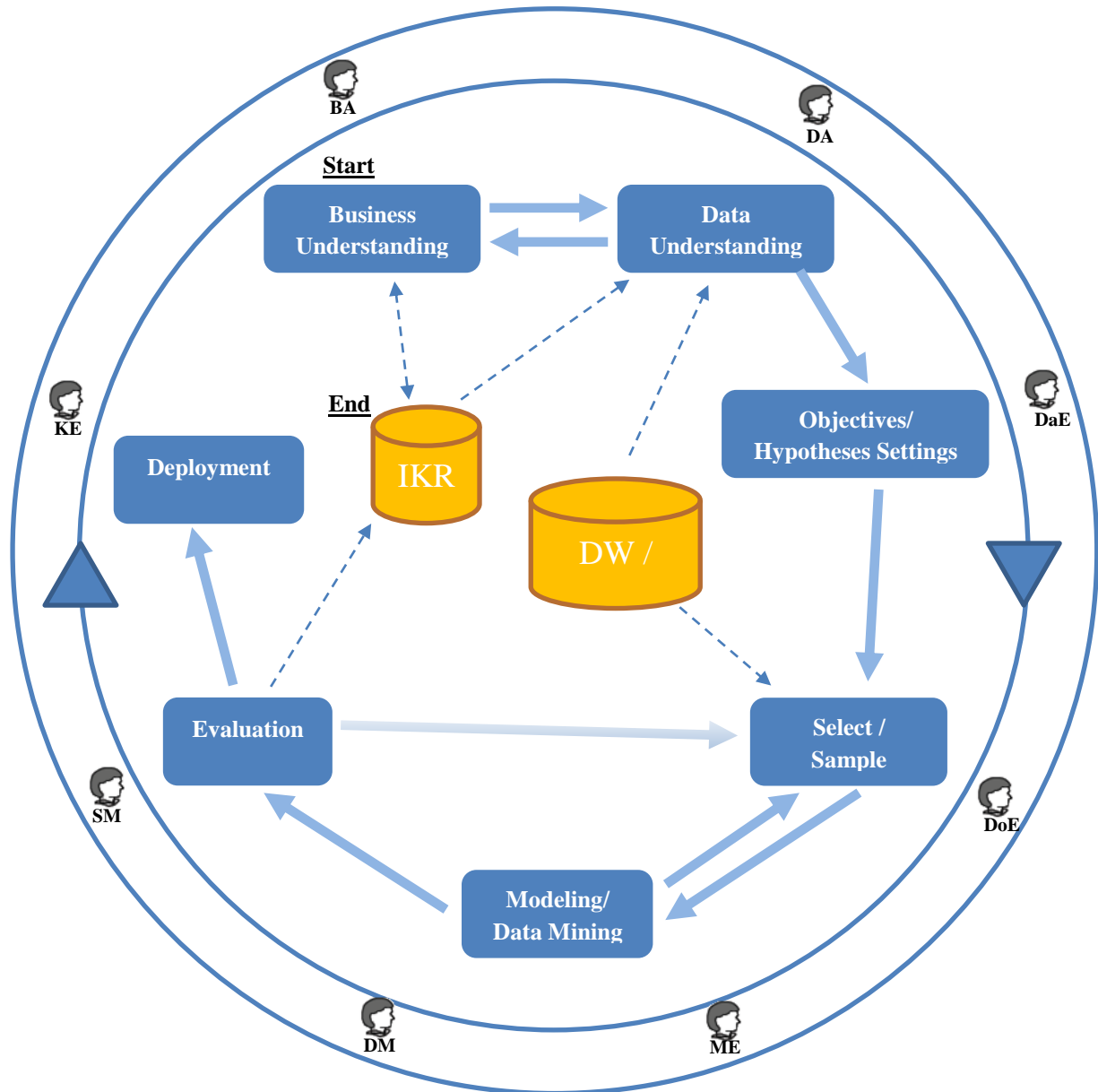
تعتمد بنية النموذج (SCRUM-BI) على جوانب ثلاثة أساسية تدعم وتعزز شمولية النموذج وتكامله، وهي: العمليات، الأفراد المشاركة بالعمل، مخازن البيانات. فنحرص في نموذجنا على: (I) تعريف العمليات التي تتضمنها دورة حياة النموذج، والأنشطة والمهام التي تتم خلال كل عملية. (II) تعريف مصادر البيانات ومخازنها، وتدقيقها خلال دورة

حياة النموذج. (III) تعريف الأفراد التي تتحمل أعباء القيام بهذه العمليات، وأدوار كل منهم خلال دورة حياة النموذج. كل ذلك بهدف ضمان نجاح مشاريع ذكاء الأعمال (BI) وتحقيق أهدافها. يتألف النموذج (SCRUM-BI) من سبع عمليات أساسية: فهم البيانات، وفهم بيئة العمل، ووضع الأهداف والفرضيات، واختيار عينات البيانات، والنمذجة والتنقيب البيانات، والتقييم، والنشر. يتم تخزين البيانات والمعلومات والمعارف في مخزنيّ بيانات: مستودع البيانات وأسواقها (DW/DMS)، ومخزن المعرفة والمعلومات (IKR). ويتضمن النموذج مجموعة من الأفراد الخبيرة والموارد البشرية المتضمنة خلال دورة حياته. يمرّ المشروع بعدة دورات يتمّ تحديدها منذ بدايته، كما يحوي مسارات داخلية تسمح بالعودة لأحد العمليات السابقة، وتصحيح مسار المشروع دون الحاجة لانتظار دورات المشروع اللاحقة. سنقوم فيما يلي بتفصيل الجوانب الأساسية في النموذج (SCRUM-BI):

٣-١. العمليات Processes

تتكون دورة حياة النموذج (SCRUM-BI) من سبع عمليات متتالية أساسية، تمّ اختيارها من عدة نماذج سابقة: (DMLC, 2009)، (CRISP-DM, 2008). وذلك للاستفادة من نقاط قوة كل نموذج وتلافي مواضع القصر فيها، والتي تم توضيحها في الدراسة المرجعية المقدمة في هذا البحث. فقد تمّ اختيار هذه العمليات بحيث تكون شاملة وتفصيلية، وذات معنى تقوم على حجم عمل جيّد، ولا تشكل عبئاً على إدارة دورة حياة النموذج بنفس الوقت. إنّ خرج كل عملية يحدّد العملية اللاحقة، كما أنّ بعض العمليات تتمّ على التوازي؛ حيث يتمّ العمل خلالها ذهاباً وإياباً حتى تحقيق الهدف من كل عملية، فلا يمكن الفصل بين عمليتي اختيار البيانات من جهة والنمذجة والتنقيب في البيانات من جهة أخرى على سبيل المثال، حيث يمكن تطبيق نفس خوارزمية التنقيب في البيانات وتجريبها على أكثر من عينة للوصول للنتائج المرجوة، وذات الأمر بالنسبة للعمليتين: فهم البيانات وفهم بيئة العمل؛ ففهم بيئة العمل يساعد في فهم البيانات وما تتضمنه مستودع البيانات وأسواقها (DW/DMS)، كما أنّ البيانات تعطي صورة عن بيئة العمل. الأسهم المبينة في الشكل (٢١) توضح المسار الرئيسي للحلّ المقدم في هذا العمل، فيما توضح الحلقة الخارجية أنّ النموذج (SCRUM-BI) يمرّ بعدد من الدورات، يتمّ تحديد عددها والزمن المحدد لكل منها (Timebox) منذ بداية المشروع، والذي يتفاوت بين ٨ أسابيع و ١٢ أسبوع كحد أقصى، فالمعرفة المستخرجة في الدورات البدائية، والخبرة المتراكمة خلال العمليات في هذه الدورات، تسهم في فهم أكبر لبيئة العمل والبيانات، وتوضيح المشاكل فيها، مما يسهم في وضع الفرضيات وتحديد الأهداف بشكل أفضل، كما يسهم في تحديد أساليب النمذجة وخوارزميات التنقيب في البيانات، بالاستفادة من الخبرات المتراكمة خلال الدورات السابقة.

تتضمن كل عملية مجموعة من الأحداث والمفاهيم التي تمّ اعتمادها بالاستفادة من المنهجية سكروم، حيث يتمّ تنفيذ كل عملية خلال شوط واحد كما وضحنا في الفصل السابق من هذا البحث، وسيتم في بداية كل عملية عقد اجتماع (التخطيط للعملية ويمتدّ هذا الاجتماع لثماني ساعات كحد أقصى) بين أفراد فريق العمل المسؤول عن العملية للتخطيط للعملية، ينتج عنه خطة العملية وتتضمن: مجموعة المتطلبات والأعمال والوظائف التي سيتم إنجازها خلال العملية، وحجم العمل اللازم لإنجازها، والتقدير الزمنية لها، والموارد المتوفرة، والهدف النهائي للعملية، والزمن اللازم لإنجاز العملية، ويتراوح بين أسبوع وثلاث أسابيع كحد أقصى. كما سيتم عقد اجتماع آخر (استعراض العملية ويمتدّ هذا الاجتماع لثماني ساعات كحد أقصى) في نهاية الشوط بين أفراد فريق العمل من جهة، وأصحاب المصلحة بالمشروع (متخذو القرار، الزبائن، ...) من جهة أخرى، لاستعراض خرج العملية، وتحديد ما تمّ إنجازه، والمشاكل التي تمّ مواجهتها، والتحسينات الممكنة خلال العمليات والدورات اللاحقة، والتحضير للعملية التالية. بالإضافة للاجتماعات اليومية قصيرة المدة (لا تتجاوز ربع ساعة)، يقوم فريق العمل خلالها بمزامنة جميع الأنشطة التي يقومون بها، والتخطيط للعمل المراد القيام به خلال اليوم التالي، والتعديل على خطة العملية: بحذف ما يروونه غير ضروري، أو إضافة عناصر أخرى بحسب التغييرات الطارئة على المتطلبات، بعد إطلاع أصحاب المصلحة عليها. الجدير بالذكر أنّ خبير المنهجية يعتبر الشخص المسؤول عن تطبيق هذه الأحداث والاجتماعات، وتعريف فريق العمل بما يجب القيام به في كل مرحلة، وإدارة التواصل بين فريق العمل والأفراد الخارجيين أصحاب المصلحة (مدراء، زبائن، ...).



الشكل (٢١) بنية النموذج SCRUM-BI

سنعرض فيما يلي قائمة العمليات التي يتكون منها النموذج (SCRUM-BI)، والموضحة بالشكل (٢١):

٣-١-١. فهم بيئة العمل (Business Understanding)

يمكن اعتبارها نقطة انطلاق النموذج (SCRUM-BI) وواحدة من أهم العمليات فيه، حيث تركز على فهم أهداف ومتطلبات المشروع من وجهة نظر بيئة العمل، بالشكل الذي سيساعد في تعريف أهداف وفرضيات المشروع لاحقاً. فعلى جميع أفراد فريق العمل أن يكون على دراية تامة حول كيفية سير العمل، والقواعد والشروط الخاصة ببيئة العمل، والأهداف الخاصة بأصحاب العمل، وما يتوقعونه في نهاية المشروع. يتم خلال هذه المرحلة العديد من الأنشطة الكفيلة بتحقيق هدف العملية، أهمها:

- **تحديد الأهداف الخاصة ببيئة العمل:** فالهّم الأساسي لمحلل الأعمال في هذه المرحلة، فهم ما يرغب صاحب العمل الحصول عليه في نهاية المشروع، والكشف عن عوامل مهمة في بيئة العمل من الممكن أن تؤثر على المشروع ككل، بحيث يمكن الحفاظ على التوازن بين رغبات وأهداف أصحاب المصلحة بالمشروع من جهة، وبين شروط وقيود بيئة العمل من جهة أخرى.
- **تقدير الوضع الحالي لبيئة العمل:** بهدف تحديد جميع المتطلبات والفرضيات والشروط والتقديرية والموارد والمخاطر وغيرها من عوامل بيئة العمل، وبناء تقارير المنافع والتكاليف، بما يسهم بوضع أهداف المشروع ووضع خطة العمل.

يتم تخزين خرج هذه العملية في مخزن المعرفة والمعلومات (IKR)، للسماح بمشاركتها مع أفراد العمل بما يضمن تحقيق أهداف المشروع، ويتضمن هذا الخرج المعلومات التي تمّ الحصول عليها في هذه العملية من: الفرضيات والقواعد الخاصة ببيئة العمل، وأهداف العمل، والتقديرية المتعلقة ببيئة العمل، وتقارير التكاليف والمنافع. يسمح الفصل بين المعلومات الخاصة ببيئة العمل والمعلومات الخاصة بالبيانات (meta data)، والتي يتم تخزينها في مستودع البيانات وأسواقها (DW\DMs)، بالحفاظ على اتساق وتكامل هذه المعلومات ويقلل من التكرار.

يتحمل محلل الأعمال العبء الأكبر في هذه العملية، ويلعب كل من: محلل البيانات وخبير مجال العمل والمدير الاستراتيجي دوراً مساعداً، إلى جانب خبير المنهجية الذي يشرف على نجاح الأحداث والاجتماعات في هذه العملية.

٣-١-٢. فهم البيانات (Data Understanding)

تعتبر طبقة البيانات والفهم الصحيح لها، أساساً لنجاح تطبيقات ذكاء الأعمال (BI) والتنقيب في المعطيات، فتعتبر العملية الأساسية الثانية إلى جانب عملية فهم بيئة العمل في النموذج. وهي تتمّ جنباً إلى جنب مع عملية فهم بيئة

العمل، حيث يساعد فهم بيئة العمل وتحديد الفرضيات والشروط والقيود الخاصة به في فهم أفضل للبيانات، والعكس بالعكس إنّ البيانات تعطي صورة أفضل عن منظور العمل. تبدأ هذه العملية باستكشاف البيانات في مستودع البيانات وأسواقها (DW\DMS)، للقيام بجميع المهام والأنشطة اللازمة لجمع المعلومات الخاصة بالبيانات وتوصيفها، ومشاركة هذه المعلومات مع جميع أفراد فريق العمل ليفهموا البيانات التي يتعاملون معها، بالشكل الذي يساعد في تحديد درجة جودة البيانات، والمشاكل المتعلقة بها، وصياغة الفرضيات في المراحل اللاحقة من دورة حياة النموذج. من أهم المهام والأنشطة اللازمة في هذه العملية:

- **توصيف البيانات:** وتتضمن هذه المهمة تسجيل حجوم البيانات، وأنماطها، وعدد الجداول، وعدد السجلات في كل جدول، وعدد الحقول في كل جدول، وأنماط الحقول، وتسجيل هذه المعلومات في مخزن البيانات المترفعة الخاص بمستودع البيانات وأسواقها (DW\DMS). بما يسهم في الفهم العام للبيانات ودلالاتها، بالشكل الذي يخفف الأعباء في عمليات اختيار البيانات وتمذجتها في المراحل اللاحقة من دورة حياة النموذج.
- **استكشاف البيانات:** بهدف فهم أكبر لمحتويات البيانات، والقواعد الخاصة بها، وإظهار النتائج من خلال مجموعة من المخططات والرسوم البيانية، كتحديد نمط المخطط الخاص بمستودع البيانات (DW)، وجداول الحقيقة (Fact Tables)، والأبعاد، والقيم المفقودة، والقيم المتطرفة. كل ذلك يسهم في تحديد أفضل لأهداف عمليات التنقيب في البيانات، واختيار البيانات، واختيار خوارزميات التنقيب في البيانات.
- **تحديد درجة جودة البيانات:** لتحديد ما يجب القيام به في عمليات اختيار البيانات اللاحقة، حيث يتم في هذه المهمة معالجة مسائل مثل: هل البيانات كاملة وتغطي جميع الحالات المطلوبة؟ هل هي صحيحة أو تحوي أخطاء؟ وهل يمكن تصحيح هذه الأخطاء أو التخلص منها؟ هل هناك قيم مفقودة وكيف يتم التعامل معها؟ لیتم وضع حلول لهذه المشاكل في نهاية المهمة، بالاعتماد على المعلومات المستخرجة من فهم البيانات من جهة، وفهم بيئة العمل من جهة أخرى.

يتم تخزين المعلومات وتقارير كل من توصيف البيانات وجودة البيانات المستخرجة في نهاية هذه العملية في مخزن معلومات البيانات والبيانات المترفعة (Metadata Repository) الخاص بمستودع البيانات وأسواقها (DW\DMS). فيما يتحمل محلل البيانات العبء الأكبر في هذه العملية، ويلعب كل من: محلل الأعمال ومهندس البيانات دوراً مساعداً، إلى جانب خبير المنهجية الذي يشرف على نجاح الأحداث والاجتماعات في هذه العملية.

٣-١-٣. وضع الأهداف والفرضيات (Objectives & Hypotheses Setting)

تعتبر هذه العملية آخر عملية من عمليات التحضير للقيام بالنمذجة والتنقيب في البيانات، فتستخدم هذه العملية خرج العمليتين السابقتين لتحديد الفرضيات ووضع أهداف تطبيق ذكاء الأعمال، تم الإشارة لهذه العملية بدايةً من قبل [43]، ليتم التأكيد على أهمية وضع الأهداف والفرضيات بالتقاطع مع فهم بيئة العمل وفهم البيانات، مما "يعطي رؤية أفضل لبيئة العمل، ويضيء على نقاط أخرى فيها" [22]. بالمقابل لم يحوي نموذج (CRISP-DM, 2008) هذه العملية بشكل واضح، وإنما تاه بين أهداف العمل وهي مهمة يتم إنجازها خلال عملية فهم بيئة العمل، وأهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في المعطيات. فأهداف التنقيب في المعطيات تحقق أهداف العمل في النهاية ولكن لا يمكن أن تكون هي نفسها. فعلى سبيل المثال:

- أهداف العمل لشركة ما: زيادة عدد الزبائن.
- أهداف تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات: توقع عدد الزبائن التي ستتعامل مع الشركة نتيجة لطرح عرض جديد، بناءً لتحليل البيانات الخاصة بسلوك الزبائن خلال السنوات الأربع الأخيرة عند طرح عروض معينة من جهة، والمعلومات الخاصة بالزبائن (العمر، الوظيفة، ...) من جهة أخرى.
- الفرضيات الخاصة بتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات: الحصول على عدد الزبائن المتوقع تعامله مع الشركة عند طرح عرض جديد، يسهم في عملية اتخاذ القرار الخاصة بطرح العرض أو عدمه، وبالتالي معرفة زيادة عدد الزبائن المتوقع.

يتم في هذه المرحلة وضع أهداف الدورة الحالية من المشروع، وخطة تحقيقها، وتحديد عوامل نجاحها، وتصميم التقارير ولوحات التحكم المتوقع الحصول عليها في نهاية هذه الدورة من المشروع، واستعراضها مع أصحاب المصلحة في نهاية هذه العملية، مما يسمح بالحصول على ردود أفعالهم في وقت مبكر من عمر الدورة الحالية للمشروع. ومن المهام والأنشطة الأساسية في هذه العملية:

- تحديد أهداف الدورة الحالية من المشروع: بهدف توصيف المخرجات المرجوة من هذه الدورة من المشروع بشكل دقيق، والتقارير ولوحات التحكم المتوقع الحصول عليها.
- تحديد معايير نجاح المشروع: كتحديد مستوى دقة التوقع أو التنبؤ، وتقدير النتائج بالشكل الذي يساعد في استبعاد القيم المستحيلة وبالتالي تقليص الأخطاء في المراحل اللاحقة من دورة حياة النموذج.

تتم هذه العملية بالتعاون بين: المدير الاستراتيجي، ومحلل البيانات، ومهندس البيانات، والخبير بمجال العمل، ومحلل الأعمال، وخبير المنهجية. بالإضافة للاعتماد على المعلومات والمعارف المخزنة في مستودع البيانات وأسواقها (DW\DMS) ومخزن المعلومات والمعرفة (IKR)، والتي تم الحصول عليها في العمليتين السابقتين.

٣-١-٤. اختيار عينات البيانات (Select/Sample)

يتم في هذه العملية اختيار البيانات المستهدفة، والتي سيتم تنفيذ خوارزميات التنقيب في البيانات عليها. فالهدف من هذه العملية اختيار عينة البيانات من مصادرها الضخمة والكبيرة الحجم كمستودع البيانات (DW)، للحد من حجم البيانات، كتطبيق تقنيات التقليل (discretization) على سبيل المثال. فتم هذه العملية جنباً إلى جنب مع عملية النمذجة والتنقيب في البيانات، بحيث يتم اختبار خوارزميات التنقيب في البيانات على أكثر من عينة، حتى الحصول على عينة البيانات الصحيحة والملائمة، والتي سيتم استخدامها في عمليات النمذجة والتنقيب في البيانات في العمليات اللاحقة، وعملية اختيار العينة هذه لا تعود فقط لأهداف عمليات التنقيب في البيانات بل تخضع لشروط تقنية أخرى متعلقة بجودة البيانات ونمطها. تتضمن عملية اختيار عينة البيانات جميع المهام والأنشطة اللازمة لاختيار وبناء وتحويل البيانات التي سيتم استخدامها في عمليات النمذجة والتنقيب في البيانات، ومن أهم هذه المهام والأنشطة:

- **اختيار عينة البيانات:** يتم اختيار عينة البيانات بناءً على مجموعة من المعايير مرتبطة بأهداف عملية التنقيب في البيانات، وجودة البيانات، وشروط تقنية مثل حجم عينة البيانات ونمطها، من خلال تطبيق تقنيات يدوية وآلية لاختيار السجلات والحقول والواصفات والمميزات المطلوبة.
- **معالجة عينة البيانات:** على الرغم من اختيار البيانات من مستودع البيانات وأسواقها (DW\DMS)، إلا أنّها ما تزال بحاجة لمجموعة من الخطوات لتتلاءم وتتوافق مع الاحتياجات الخاصة بعمليات النمذجة والتنقيب في البيانات. ومن هذه الخطوات: تنظيف البيانات من القيم المتطرفة التي تقع خارج حدود قيم معظم بيانات العينة، أو القيم المفقودة، أو القيم الخاطئة، أو القيم غير الرقمية عندما نتوقع قيمة رقمية. كل ذلك بهدف تحسين جودة البيانات وبالتالي تحسين نتائج عمليات النمذجة والتنقيب في البيانات.
- **تحويل عينة البيانات:** بعد بناء عينة البيانات وتنظيفها، لا بد من القيام بمجموعة من الخطوات للتأكد من تكاملها وانسجامها، لتتوافق أخيراً مع جميع المتطلبات والقيود الخاصة بعملية النمذجة والتنقيب في البيانات. ويتضمن تحويل عينة البيانات: عمليات تقليل البيانات والإسقاط (Reduction and Projection).

ينتج عن هذه العملية عينات البيانات التي سيتم تطبيق خوارزميات التنقيب في البيانات عليها وتوصيفها. ويتعاون كل من مهندس البيانات والمنقب في البيانات على تحديد عينة البيانات هذه بشكل أساسي، وبمساعدة محلل البيانات وإشراف خبير المنهجية.

٣-١-٥. النمذجة والتنقيب في البيانات (Modeling/Data Mining)

يمكن أن يحدث التنقيب في البيانات فرقاً لأي صناعة أو مؤسسة في العالم، فيمكن التنقيب في البيانات واستخدام النتائج ليس فقط لمعرفة ما يحتاجه الزبون، بل للتوقع والتنبؤ بما سيحتاجه في المستقبل. يعتبر التنقيب في البيانات مكوناً أساسياً لتطبيقات ذكاء الأعمال، فالهدف الرئيسي لهذه العملية إجراء مهام التحليل والتنقيب في البيانات، بالاعتماد على نماذج مختلفة من نماذج التنقيب في البيانات: الشبكات العصبونية، وشجرات القرار، والتعليل المعتمد على الذاكرة، وغيرها من النماذج. كما يوجد العديد من خوارزميات التنقيب في البيانات، واستخدام إحدى هذه الخوارزميات يختلف بحسب أهداف المشروع، وبعض هذه الخوارزميات والتقنيات قد يحتاج صيغة معينة من البيانات، مما يتطلب العودة لاختيار البيانات أكثر من مرة، حتى الحصول على عينة البيانات المناسبة. وتعتبر الخوارزميات التالية من أشهر خوارزميات التنقيب في البيانات:

- **التصنيف (classification):** تجميع البيانات بناءً لمجموعة من الخصائص والمميزات المعرفة مسبقاً، كتصنيف الزبائن على أساس المهنة والجنس على سبيل المثال.
- **العقدة (clustering):** تمييز الصفات والمميزات الأساسية التي تشترك فيها عينة من البيانات، فتختلف عن التصنيف بأنّ المميزات والخصائص غير معرفة مسبقاً، ويتم اكتشافها خلال تطبيق الخوارزمية على عينة البيانات.
- **التجميع (association):** الكشف عن العلاقات والارتباطات بين البيانات، والتي لا تظهر في مستودع البيانات (DW) العادي، ومن الأمثلة الشائعة في هذا الموضوع: الزبون الذي يشتري البيض، سوف يشتري حليب بنسبة ٨٠٪.
- **التسلسل (sequencing):** تشبه خوارزمية التجميع ولكن تتمّ خلال فترة زمنية طويلة، وليس على أحداث محددة فقط.

يتم خلال هذه العملية استخراج المعلومات والمعرفة من عينات البيانات المعتمدة، ويعتبر المنقب في البيانات المسؤول الأساسي عن اختيار نماذج وخوارزميات التنقيب في البيانات، واختباره على عينات البيانات التي تم اختيارها، ويساعده خبير مجال العمل في التأكد من تحقيق هذه العملية لأهداف بيئة العمل، والتزامها بالفرضيات والقيود الخاصة بها، ويشرف خبير المنهجية على الأحداث والاجتماعات في هذه العملية.

٣-١-٦. التقييم (Evaluation)

هي واحدة من أهم العمليات في الحل المقترح، حيث يتمّ خلالها التحقق من صحة النموذج المقترح أو خوارزميات التنقيب في البيانات المستخدمة، والنتائج التي تم الحصول عليها قبل نشرها في العملية اللاحقة، وضمان تحقيق أهداف المشروع، والتي تمّ وضعها وتحديدتها في مرحلة سابقة. فالجهد الأكبر في هذه العملية ينصب على تقييم النموذج الذي قمنا ببنائه، وتقييم النتائج التي تمّ الحصول عليها، أو المقارنة بين عدة خوارزميات للتنقيب في البيانات ونتائجها، لاختيار النتائج التي تحقق الأهداف المطلوبة واستخدامها ونشرها لاحقاً، وتحديد الخطوات اللاحقة من العمل. وتلعب الاجتماعات مع أصحاب المصلحة (مدراء، زبائن، ...) الدور الأساسي في هذه العملية. ويمكن اعتبار الأنشطة والمهام التالية أساسية في عملية التقييم:

- **تقييم نتائج عملية النمذجة والتنقيب في البيانات:** حيث ينتج عنها العديد من التقارير والأنماط، وينصب اهتمامنا في عملية التقييم على التحقق من صحتها وتحقيقها لأهداف المشروع، على اعتبار أننا قمنا خلال عملية وضع الأهداف والفرضيات في مرحلة سابقة من دورة حياة النموذج (SCRUM-BI): بتحديد أهداف المشروع، والقيود المتعلقة به من جهة منظور بيئة العمل، والشروط الخاصة بالبيانات، ووضع التصميم الخاصة بالتقارير ولوحات التحكم، ومراجعتها مع أصحاب المصلحة (المدراء، الزبائن، ...).
- **تحديد الخطوات اللاحقة:** فبعد مراجعة جميع الأعمال السابقة، وتقييم ما تمّ التوصل إليه، يُتخذ القرار بنشر النتائج التي تمّ توصل لها، أو العودة لملاحظات داخلية لاختبار النماذج وخوارزميات التنقيب في البيانات على عينات أخرى من البيانات.

ينبغي في نهاية هذه العملية التوصل لقرار حول النتائج التي سيتم استخدامها ونشرها، بالمقابل تُحزّن نتائج النمذجة والتنقيب في البيانات والتي تمّ التحقق من صحتها في مخازن المعرفة والمعلومات (IKR)، حتى وإن لم يتم استخدامها أو نشرها، وذلك للاستفادة منها في الدورات اللاحقة من المشروع أو في مشاريع أخرى. يتعاون كل من الخبير في مجال العمل، ومحلل الأعمال، والمدير الاستراتيجي، للتأكد من صحة النتائج وتحقيقها الغرض المطلوب، فيما يعتبر مهندس المعرفة مسؤولاً عن تخزين المعارف والمعلومات الجديدة، ويشرف خبير المنهجية على الأحداث والاجتماعات في هذه العملية.

٣-١-٧. النشر (Deployment)

هي آخر عملية لكل دورة من دورات النموذج (SCRUM-BI)، فالنتائج والمعرفة المكتشفة بحاجة لتنظيم وإدارة، وإيجاد طرق العرض والإظهار بالشكل الذي يمكن المستخدم النهائي من فهمها واستخدامها، وذلك بالاعتماد على التقارير ولوحات التحكم التي تمّ تصميمها ومناقشتها مع أصحاب المصلحة (مدراء، زبائن، ...) سابقاً خلال عملية

وضع الأهداف والفرضيات. كما تتيح هذه العملية مشاركة المعارف والخبرات والمعلومات الجديدة التي تم اكتشافها بين أفراد فريق العمل للاستفادة منها خلال الدورات اللاحقة من المشروع. فيزود نشر نتائج التنقيب في البيانات الموظفين بالمعلومات التي يحتاجون لها، بالصيغة التي تمكنهم من استخدامها، حيثما يحتاجونها، وعندما يحتاجونها.

ليتمّ أخيراً تقييم نتائج الدورة الحالية للمشروع، وما تمّ إنجازه، والإعداد للدورة القادمة، بالتعاون بين جميع أفراد فريق العمل، فيما يشرف المدير الاستراتيجي وخبير مجال العمل على مهام نشر المعارف المكتشفة وإظهارها، والتأكد من فهمها بشكل جيد من قبل الجميع، ويشرف خبير المنهجية على الأحداث والاجتماعات في هذه العملية.

٣-٢. مخازن البيانات Data Repositories

تعتبر مخازن البيانات من المقومات الأساسية لتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، فيجب تعريف دخل أي نموذج لاكتشاف المعرفة في البيانات (KDD) وخرجه بشكل واضح. العديد من النماذج السابقة لم توضح طبيعة مصادر البيانات وبينيتها؛ فالبعض اعتبر أنّ أي مصدر للبيانات يمكن اعتماده لإجراء عمليات التحليل والتنقيب في البيانات (Gottgorty, 2010 [38]; Lee and Kerschberg, 1998 [57]; Fayyad et al., 1996 [2])، فيما لم يتطرق آخرون لهذه النقطة أبداً خلال توصيف نماذجهم ودورة حياتها (CRISP-DM, 2008 [41]; Collier et al, 1998 [43]; SEMMA, 1997 [40])، بالمقابل أشار العديد من الباحثين لضرورة اعتماد مستودع البيانات وأسواقها (DW\DMs) كدخل لاكتشاف المعرفة في البيانات (KDD) (Alnoukari, 2012 [7]; Hofmann, 2009 [10]; Kopanakis and Theodoulidis, 1999; Feldens, 1998 [36]; Ganesh et al., 1996). أما بالنسبة لتخزين المعارف الجديدة فيعتبر (Lee and Kerschberg, 1998) من أوائل من استخدم مخازن المعرفة والمعلومات في اكتشاف المعرفة في البيانات (KDD)، لتخزين المعرفة والمعلومات الجديدة، ليتم لاحقاً اعتماد هذه المخازن من قبل العديد من النماذج الأخرى (Alnoukari M, 2012; Hofmann M, 2009).

يستخدم النموذج (SCRUM-BI) مستودع البيانات وأسواقها (DW/DMs) كمصدر أساسي للبيانات، ومكان تخزينها، وتخزين البيانات المترقعة حولها، فيما يستخدم مخزن المعرفة والمعلومات (IKR) كمكان تخزين المعارف والمعلومات الجديدة، والتي تم اكتشافها والحصول عليها خلال دورة حياة النموذج.

على الرغم من أنّ أي مصدر للبيانات منظم أو شبه منظم، يساعد بالقيام بعمليات التحليل والتنقيب في البيانات بشكل أفضل، إلا أنّ معالجة البيانات، وتنظيمها، ووضعها في بنية هيكلية جيدة يضمن جودة النتائج المتوقعة ويعززها،

وهذا ما يتوافق مع بنية مستودع البيانات وأسواقها (DW/DMS)، التي تشكل مصدراً للبيانات لا مفرّ منه في مشاريع ذكاء الأعمال (BI) والتنقيب في البيانات [22].

يستخدم النموذج (SCRUM-BI) مستودع البيانات وأسواقها (DW/DMS) كمصدر أساسي للبيانات، والتي سيتم عليها عمليات التحليل والتنقيب في البيانات. فيزود مستودع البيانات (DW) بالبيانات من مختلف الأنظمة التشغيلية، ويتم نقل البيانات لها باستخدام أدوات (ETL)، من خلال استخراج البيانات من مصادرها، وتحويلها، وتحميلها في مستودع البيانات (DW). ولا بد هنا من التمييز بين مستودع البيانات (DW) وهو مخزن فيزيائي مركزي، حيث يتم تخزين البيانات المجمعة من مختلف الأنظمة على مستوى المؤسسة ككل، بحيث تكون نظيفة، ومتكاملة، ومهيكلية، ومصاغة بشكل معياري موحّد، وبين أسواق البيانات وهي عبارة عن جزء من مستودع البيانات (DW)، لتخزين عينات البيانات ذات الصلة بعمليات التحليل والتنقيب في البيانات. إلى جانب مستودع البيانات وأسواقها (DW/DMS) يوجد مخزن البيانات المترقّعة (Metadata Repository)، التي تعدّ حجر الزاوية لتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، ومن دونها يمكن أن يزداد عبء العمل في مشاريع ذكاء الأعمال (BI) والتنقيب في البيانات بشكل كبير، فتساعد البيانات المترقّعة في فهم البيانات، وتخزين دلالات الحقول والسمات: أنماطها، وقيمها، والقيود عليها. يحوي مخزن البيانات المترقّعة معلومات حول البيانات الموجودة في مستودع البيانات وأسواقها (DW/DMS)؛ كيف، ولماذا، وأين يمكن العثور عليها، وتخزينها، واسترجاعها، واستخدامها. توصف البيانات المترقّعة الموجودة في مستودع البيانات وأسواقها (DW/DMS) على كافة المستويات: الدلالية، والإحصائية، والهيكليّة، والفيزيائية، من أجل دعم عمليات اختيار البيانات، وتدقيقها، وتطبيق خوارزميات التنقيب في البيانات، وتفسير النتائج. وقد عرّف Pant مخزن البيانات المترقّعة بأنه "مخزن جميع المعلومات حول المصدر، والوجهة، والتحويلات، والجلسات، والمخططات، وتدفعات العمل، وشروط بيئة العمل". والجدير بالذكر أخيراً أنّ مخزن البيانات المترقّعة يوجد فيزيائياً إلى جانب مستودع البيانات وأسواقها (DW/DMS) نفسه.

تعتبر قضية تخزين المعارف والمعلومات الجديدة التي تم اكتشافها خلال دورة حياة النموذج إحدى أهم القضايا في بنية النموذج (SCRUM-BI)، فيتم تخزين المعرفة الصريحة في مخزن المعرفة والمعلومات (IKR)، وهو عبارة عن قاعدة بيانات علائقية، تمكن من تخزين المعارف الجديدة بشكل فعّال. ترتبط عملية التخزين هذه بمجموعة من العمليات الأخرى، أهمها: الاستحواذ على المعرفة الجديدة، وصياغتها بالشكل الذي يساعد في تخزينها، وتنظيم المعرفة ضمن مخزنها، وإدارة الوصول لهذه المعارف ضمن المخزن، واسترجاع المعرفة من المخزن. ونورد فيما يلي لمحة عن أهم أنماط المعارف في المؤسسات:

- المعرفة الخارجية: مثل الذكاء التنافسي (competitive intelligence)، والتي تحتاج للتوضيح والترجمة.

- **المعرفة الدّاخلية:** كتقارير الأبحاث، والعروض التقديمية، والمواد التسويقية، والتي بالمعظم يمكن القول بأنّها معرفة صريحة.

يُستخدم مخزن المعرفة والمعلومات (IKR) كدخل لعمليات فهم بيئة العمل ووضع الأهداف والفرضيات، وذلك للاستفادة من المعارف الموجودة فيه في تلك العمليات. إلى جانب استخدام مخزن المعرفة والمعلومات (IKR) كخرج للنموذج (SCRUM-BI)، لتخزين المعارف والمعلومات المكتشفة خلال دورة حياة النموذج، فيحوي مخزن المعرفة والمعلومات (IKR) القواعد المتعلقة ببيئة العمل ونتائج عمليات تحليل البيانات والتنقيب فيها، والمصاغة بشكل يمكن من تخزينها، وتنظيمها، واسترجاعها.

٣-٣. الموارد البشرية Human Resources

تلعب الموارد البشرية دوراً هاماً في أي عملية لاكتشاف المعرفة وتنظيمها. تجاهلت معظم النماذج السابقة لاكتشاف المعرفة في البيانات (KDD) أهمية تضمين الموارد البشرية في نماذجهم، فيما قدم (Ganesh et al., 1996) محلل البيانات والمنقب في البيانات، وقدم (Lee et al., 1998) خبير مجال العمل، وقدم (Kopanakis et al., 1999) مهندس المعرفة، ويعتبر (Hofmann, 2009) من أوائل الباحثين الذين شدّدوا على أهمية إشراك الموارد البشرية في دورة حياة اكتشاف المعرفة في البيانات (KDD)، وقدم عدد كبير من الأفراد ذوي المهارات المختلفة في نمودجه DMLC، وقد تضمنت: مدير المشروع، ومحلل الأعمال، ومحلل البيانات، ومهندس البيانات، والمنقب في البيانات، وخبير مجال العمل، ومهندس المعرفة، والمدير الاستراتيجي. بالمقابل ضمّن (Alnoukari, 2012) الخبرات السابقة نفسها في نمودجه ASD-BI، ولكنّه قدم محلل البيانات ومهندسها كدور واحد في نمودجه؛ كونها تتطلب نفس الخبرات والإمكانيات، كما استبعد مدير المشروع على اعتبار أنّ الموارد البشرية في اكتشاف المعرفة في البيانات (KDD) تختلف عن الموارد البشرية في عملية إدارة المشاريع، والتي تتطلب خبرات ومهارات أخرى بما في ذلك: مدير المشروع، ومطور التطبيقات (Developer)، ومختبر التطبيقات (Tester)، وغيرهم.

يتضمن النموذج (SCRUM-BI) المقدم في هذا البحث الموارد البشرية التالية: (I) خبير المنهجية. (II) محلل الأعمال. (III) محلل البيانات. (IV) مهندس البيانات. (V) المنقب في البيانات. (VI) خبير مجال العمل. (VII) مهندس المعرفة. (VIII) المدير الاستراتيجي. سنقوم بتوصيف مهام كل منهم فيما يلي، محاولين توضيح سبب اعتمادهم في نموذجنا (SCRUM-BI):

٣-٣-١. خبير المنهجية

يعتبر المسؤول عن إدارة الاجتماعات والأحداث خلال العمليات التي تتضمنها دورة حياة النموذج، وتوضيح أهميتها وأهدافها لأفراد فريق العمل، وضمان القيام بها بشكل سليم، والالتزام بالقواعد الخاصة بها، وإدارة العلاقات والتفاعلات بين أفراد فريق العمل. كما يساعد الأشخاص خارج المشروع بفهم كيفية التواصل والتفاعل مع فريق العمل. فيلعب خبير المنهجية دور المشرف في جميع العمليات التي تتضمنها دورة حياة النموذج.

٣-٣-٢. محلل الأعمال

يعتبر مسؤولاً عن فهم الجوانب المتعلقة ببيئة العمل من جهة، كما يلعب دور أساسي في عملية وضع الأهداف وتحديد الفرضيات بالتعاون مع محلل البيانات، وخبير مجال العمل، والمدير الاستراتيجي. بالإضافة لدور محلل الأعمال المساعد في عملية تقييم نتائج النمذجة والتنقيب في البيانات، وأثرها على العمل حالياً أو في المستقبل. ينبغي أن يتمتع محلل الأعمال بمهارات في تحليل المشاكل والمتطلبات المرتبطة ببيئة العمل، والمستقلة عن التكنولوجيا.

٣-٣-٣. محلل البيانات

يقوم بالوصول لمستودع البيانات وأسواقها (DW/DMS)، ويحلل محتواها، وبنيتها، ويقوم بتوصيفها، وتحديد درجة جودتها، ويُنشأ مخزن البيانات المترقعة (Metadata Repository). محلل البيانات هو المسؤول عن عملية فهم البيانات، ويلعب دوراً رئيسياً في عملية وضع الأهداف وتحديد الفرضيات بالتعاون مع محلل الأعمال، خبير المجال، والمدير الاستراتيجي، كما يلعب دوراً مساعداً في عملية اختيار عينات البيانات التي سيجري عليها عمليات النمذجة والتنقيب في البيانات.

٣-٣-٤. مهندس البيانات

يملك خبرة كافية في مجال تصميم وبناء مستودع البيانات وأسواقها (DW/DMS)، تجعله مسؤولاً عن عمليات اختيار عينات البيانات، ومعالجتها، وتحويلها، وبناء أسواق البيانات (DMS) المقابلة، كما يلعب دوراً رئيسياً في عملية وضع الأهداف وتحديد الفرضيات، بالإضافة لدوره المساعد في عملية فهم البيانات.

٣-٣-٥. المنقب في البيانات

يعتبر من أهم الأدوار في اكتشاف المعرفة من البيانات، فهو المسؤول عن عمليات النمذجة وتطبيق خوارزميات التنقيب في البيانات على العينات التي قام مهندس البيانات ببنائها وتصميمها. تتطلب طبيعة مهام المنقب في البيانات خبرات في كل من: قواعد البيانات، والإحصاءات، وطرق النمذجة، وخوارزميات التنقيب في البيانات، وآليات استخدامها.

يعمل المنقب في البيانات إلى جانب مهندس البيانات لاختيار عينات البيانات المناسبة، كما يساعده خبير مجال العمل في ضمان جودة نتائج خوارزميات التنقيب في البيانات وأساليب النمذجة التي تمّ اختيارها.

٣-٣-٦. خبير مجال العمل

يلعب دور هام في معظم العمليات والوظائف خلال دورة حياة النموذج (SCRUM-BI)، وذلك للاستفادة من خبراته ومعارفه في مجال بيئة العمل والقضايا المتعلقة بها، فيلعب دوراً أساسياً في وضع الأهداف وتحديد الفرضيات الخاصة بتطبيقات ذكاء الأعمال (BI) بشكل صحيح، ويحقق أهداف بيئة العمل، ويلتزم القيود والشروط المتعلقة بها، ويساعد محلل الأعمال في فهم بيئة العمل والجوانب المتعلقة بها، كما يساعد مهندس المعرفة في عملية تقييم النتائج، بما يضمن الحصول على نتائج صحيحة وذات فائدة من جهة، ويساعده في تحديد المعارف التي يجب تخزينها في مخزن المعرفة والمعلومات (IKR) من جهة أخرى.

٣-٣-٧. مهندس المعرفة

يلعب دوراً رئيسياً في ضمان الحصول على المعرفة، وتقييمها، وتمثيلها، وتنظيمها، وتخزينها في مخزن المعرفة والمعلومات (IKR)، بشكل هيكلي مثالي للاستفادة منها في عمليات فهم بيئة العمل ووضع الأهداف وتحديد الفرضيات خلال الدورات اللاحقة من دورة حياة النموذج (SCRUM-BI)، أو لاستخدامها لاحقاً في مشاريع أخرى. يعتبر مهندس المعرفة صاحب الدور الحاسم في نجاح تطبيقات ذكاء الأعمال، والاستفادة منها في تعزيز المعرفة في المؤسسات، للمساعدة في اتخاذ القرارات المناسبة.

٣-٣-٨. المدير الاستراتيجي

يهتمّ ببلورة القضايا الاستراتيجية التي ينبغي التصدي لها، بما يضمن مساهمة تطبيقات ذكاء الأعمال (BI) في صياغة الخطط الاستراتيجية للمؤسسات وتنفيذها؛ من خلال توفير البيانات، وإجراء الدراسات حول مجال العمل وشروط المنافسة فيه، ووضع تقييمات الأداء الاستراتيجي. فيلعب المدير الاستراتيجي دوراً أساسياً في عملية تقييم النتائج وضمان توافقها مع رؤية المؤسسة وتوجهاتها الاستراتيجية، ويشرف مهام نشر المعارف المكتشفة، وإظهارها، والتأكد من فهمها بشكل جيد من قبل الجميع، بالإضافة لدوره المساعد في عميات وضع الأهداف وتحديد الفرضيات.

أخيراً وبعد تحليل كل من طبيعة الموارد البشرية وبنية العمليات التي يتضمنها النموذج (SCRUM-BI) خلال دورة حياته، يمكننا بناء مصفوفة التقابل بين العمليات والموارد البشرية في النموذج (SCRUM-BI)، والموضحة بالجدول (١٢) التالي:

الجدول (١٢) مصفوفة التقابل بين العمليات والموارد البشرية في النموذج (SCRUM-BI)

الموارد البشرية	محلل الأعمال	خبير مجال العمل	محلل البيانات	مهندس البيانات	المنقب في البيانات	مهندس المعرفة	المدير الاستراتيجي	خبير المنهجية
فهم بيئة العمل	M	A	A				A	S
فهم البيانات	A		M	A				S
وضع الفرضيات وتحديد الأهداف	M	M	M	M			A	S
اختيار عينات البيانات			A	M	M			S
النمذجة والتنقيب في البيانات		A			M			S
التقييم	A	A				M	M	S
التشر	A	S	A	A	A	M	S	S

M: Major, A: Assistant, S: Supervisor.

٤. خصائص النموذج SCRUM-BI

بعد توضيح بنية النموذج (SCRUM-BI) وأهم الجوانب التي يتضمنها، يمكننا تلخيص أهم خصائص هذا النموذج وإسهاماته في ثلاث أبعاد، وهي: (I) التكيف مع التغييرات في بيئة العمل والمتطلبات. (II) البعد المعرفي للنموذج SCRUM-BI. (III) البعد الاستراتيجي للنموذج SCRUM-BI. سنقوم فيما يلي بتفصيل كل من هذه الأبعاد:

٤-١. التكيف مع التغييرات في بيئة العمل والمتطلبات

"طورت المنهجيات الرشيقة كحل لتطوير البرمجيات في ظل بيئات العمل الغامضة وغير المؤكدة" [13]. يقوم النموذج (SCRUM-BI) على الدمج والمزاوجة بين مناهج اكتشاف المعرفة التقليدية والمنهجية الرشيقة سكروم، فالمنهجية سكروم مناسبة جداً عندما تكون المتطلبات متقلبة نتيجة لديناميكية بيئات العمل والتطور المتسارع فيها حيث من الصعب تطبيق المنهجيات التقليدية. يعرف النموذج (SCRUM-BI) مجموعة من الأحداث والاجتماعات خلال العمليات التي تتضمنها دورة حياته، فسمح الاجتماعات اليومية لأفراد فريق العمل بمراجعة العمل بشكل مستمر، وتقييمه، وحذف ما هو غير ضروري من قائمة الأعمال، وإضافة أعمال أخرى عند ورود متطلبات جديدة، وإعادة كتابة التقديرات اللازمة لهذا العمل، فيما تسمح الاجتماعات التخطيط في بداية كل عملية واجتماعات مراجعة الأعمال في نهايتها بتقدير الوضع الحالي،

وتقيّم ما تمّ إنجازه، ومراجعته مع جميع الأفراد أصحاب المصلحة بمشاركة ذكاء الأعمال (BI) والتنقيب في البيانات (فريق العمل، مدراء، زائن، ...)، مما يساعد في الحصول على ردود الأفعال في وقت مبكر من عمر هذه المشاريع، والتجاوب معها، بحيث يصبح جميع الأفراد المشتركين بالعمل وأصحاب المصلحة على علم ودراية كافية عن الخرج المتوقع في نهاية كل دورة من دورات النموذج (SCRUM-BI).

تضمين خبير المنهجية في عمليات النموذج (SCRUM-BI) يضمن القيام بتلك الاجتماعات بشكل سليم، ويضمن تحقيق أهدافها، كما يدير العلاقات بين أفراد فريق العمل، والتفاعلات فيما بينهم من جهة، ومع أصحاب المصلحة بمشاريع ذكاء الأعمال (BI) والتنقيب في البيانات (مدراء، زائن، ...) من جهة أخرى، بالشكل الذي يساعد في الحصول على ردود أفعالهم، وتوضيحها لأفراد فريق العمل وكيف يجب أن يتعاملوا معها.

٤-٢. البعد المعرفي للنموذج SCRUM-BI

يعتمد النموذج (SCRUM-BI) على خصائص المنهجية سكروم والمفاهيم والأحداث التي تعرفها هذه المنهجية، مما يسمح بمشاركة الخبرات والمعارف المتراكمة خلال العمل بين جميع أفراد فريق العمل من جهة، ومع أصحاب المصلحة بمشاريع ذكاء الأعمال (BI) (مدراء، زائن، ...) من جهة أخرى، والتي يمكن أن تحدث خلال الاجتماعات اليومية قصيرة المدة، أو اجتماعات التخطيط للعمل في بداية كل عملية، أو اجتماعات مراجعة الأعمال في نهاية كل عملية. يضمن النموذج (SCRUM-BI) من خلال تضمين مهندس المعرفة عمليات الحصول على المعرفة من نتائج تحليل البيانات وتمذجتها والتنقيب فيها، وتحويل هذه المعارف المكتشفة، وصياغتها بطريقة معيارية وهيكلية تساعد في عملية عرضها وإظهارها من جهة، وعملية تخزينها في مخازن المعرفة والمعلومات (IKR) من جهة أخرى. فتوفر مخازن المعرفة والمعلومات هذه البنية والآلية لتخزين هذه المعارف المكتشفة، ومشاركتها مع الأفراد خلال دورة حياة النموذج، والاستفادة منها في مشاريع ذكاء الأعمال (BI) والتنقيب في البيانات القادمة.

٤-٣. البعد الاستراتيجي للنموذج SCRUM-BI

يركز النموذج (SCRUM-BI) على البعد الاستراتيجي لتطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات، من خلال المساعدة في صياغة الخطط الاستراتيجية وتحقيقها. فيسهل النموذج (SCRUM-BI) في صياغة أهداف هذه التطبيقات بالشكل الذي يساهم في صياغة الخطط الاستراتيجية للمؤسسات وتحقيقها، من خلال تضمينه للمدير الاستراتيجي خلال دورة حياة النموذج (SCRUM-BI) والذي يلعب دوراً أساسياً في عملية وضع الأهداف وتحديد الفرضيات من جهة، كما يساعد في تقييم النتائج، وتحديد ما هو نافع منها. تساعد تطبيقات ذكاء الأعمال (BI) ليس

فقط في معرفة ما يريده الزبائن، بل أيضاً توقع ما سيحتاجونه في المستقبل، فتكشف نتائج تطبيقات ذكاء الأعمال (BI) عن مصادر تنافسية جديدة في بيئة العمل، كما تساعد في المرونة لتلبية التغييرات فيها.

٥. خاتمة

قدمنا في هذا الفصل النموذج (SCRUM-BI) لاكتشاف المعرفة في البيانات (KDD)، والذي يقوم على الدمج والمزاوجة بين المناهج التقليدية في اكتشاف المعرفة والمنهجية الرشيقة سكروم (SCRUM). قمنا بتوضيح كيفية ملاءمة المنهجية سكروم (SCRUM) في النموذج المقدم في هذا البحث. ثم قمنا بتفصيل أهم الجوانب المتعلقة ببنية النموذج (SCRUM-BI)؛ موضحين طبيعة العمليات خلال دورة حياة النموذج، وطبيعة مصادر البيانات ومخازنها، وآليات الوصول إليها والتخزين فيها، وطبيعة الموارد البشرية المضمنة خلالها، ودور كل منها، والتفاعلات فيما بينها خلال العمليات. كما قمنا بتوضيح خصائص هذا النموذج في ثلاثة أبعاد هي: التكيف مع التغييرات في المتطلبات، والبعد المعرفي للنموذج (SCRUM-BI)، والبعد الاستراتيجي لهذا النموذج.

الفصل الخامس

الحالة الدراسية

١. مقدمة

تنتج شركات الاتصالات كمية هائلة من البيانات بشكل دائم ومستمر، ويمكن تصنيف هذه البيانات للأصناف

التالية:

- بيانات تفصيل المكالمات: التي تصف المكالمات عبر شبكة اتصال الشركة.
- بيانات عتاد الشركة: التي تصف مكونات العتاد الصلب والعتاد البرمجي ضمن الشبكة.
- بيانات الزبون: التي تصف عملاء الشركة.

تحمل هذه البيانات معلومات قيمة من الممكن تحليلها، وتوظيفها في تخطيط العمليات والتخطيط الاستراتيجي. ولكن الحجم الهائل لهذه البيانات يجعل من الصعب -إن لم نقل من المستحيل- تحليلها يدوياً، ولذلك لا بد لأن نلجأ للتحلي الآلي من خلال تقنيات النمذجة والتنقيب في البيانات.

تواجه عمليات التنقيب في بيانات شركات الاتصالات تحديات عديدة، من أبرزها:

- الحجم الهائل: قد تحوي قواعد بيانات شركات الاتصالات عشرات ملايين السجلات، الأمر الذي يجعلها من أكثر القطاعات الصناعية والتجارية توليداً للبيانات.
- طبيعة البيانات الأولية: غالباً ما تكون غير مناسبة للتنقيب في البيانات، فعلى سبيل المثال: كل من تفاصيل المكالمات وبيانات الشركة هي بيانات مسلسلة زمنياً تمثل أحداثاً فردية. فقبل أن نقوم بعمليات التنقيب في البيانات بشكل فعال، يجب بناء ملخص مفيد لهذه البيانات بحسب مجموعة من الخصائص.

- التنبؤ بأحداث نادرة جداً: العديد من تطبيقات التنقيب في البيانات في مجال الاتصالات تشمل التنبؤ بأحداث كفشل مكون ما عبر الشبكة أو حالة احتيال.

تقدم تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات حلولاً كبيرة لمعظم المشاكل التي تواجهها شركات الاتصالات في الوصول لمصادر البيانات، وتنظيمها، وتجميعها، وتحليلها، وعرض النتائج. فتساعد هذه التطبيقات في تصنيف الزبائن وتوصيفهم بحسب بياناتهم الشخصية أو استخدامهم للشبكة، وتساعد هذه التصنيفات في تمييز وتحديد الزبائن الذين يعدّون أساسيين للشركة، بالشكل الذي يمكن الاستفادة منه لاحقاً لأهداف تسويقية.

سنقوم بتوضيح التطبيق العملي للنموذج (SCRUM-BI) من خلال حالة دراسية في شركة الاتصالات الخليوية

سيرتل. نقدم في هذا الفصل إطار عمل النظام (BI-TS: Business Intelligence for Telecommunication Services) الذي قمنا بتطويره ومراحل العمل؛ مؤضحين بيئة العمل، وطبيعة البيانات التي تم التعامل معها، والأهداف المنشودة من هذا النظام، وآليات تحليل البيانات ونمذجتها والتنقيب فيها، وإظهار أبرز النتائج من خلال مجموعة من التقارير والجداول ولوحات التحكم.

٢. الإطار العام للنظام

يستخدم جميع العاملين في شركة الاتصالات سيرتل البيانات بطريقة أو بأخرى يومياً، ولكنهم يختلفون بشكل كبير بطريقة فهمهم لها، وتعامهم معها، واعتمادهم عليها في اتخاذ القرارات. نقوم في هذه الحالة الدراسية ببناء نظام ذكاء أعمال (BI) يعتمد على البيانات الموجودة لدى شركة الاتصالات سيرتل، ويتضمن ذلك مراحل مختلفة منها: دراسة الأنظمة المعلوماتية المستخدمة، وتحديد مصادر البيانات، ودراستها، وتحديد قيمتها بالنسبة لعملية اتخاذ القرار، ليتم جمعها ومكاملتها في مستودع بيانات (DW)، والذي يمثل دخل النظام (BI-TS) الذي نقوم ببنائه. سنقوم فيما يلي بتوصيف هدف النظام (BI-TS) والوظائف الأساسية، ومدخلاته ومخرجاته:

٢-١. هدف النظام BITS

تهدف من خلال النظام (BI-TS) الذي نقوم ببنائه وتطويره لتحزيم مجموعة من الخدمات (مكالمة صوتية، رسالة نصية، رسالة وسائط متعددة، تصفح الويب) ضمن حزمة واحدة أو عدة حزم، بحيث تستقطب أكبر عدد من الزبائن وتزيد من إيرادات الشركة، ودراسة تأثيرها ببعضها البعض.

٢-٢. الوظائف الأساسية للنظام BI-TS

يمكن تلخيص الوظائف الأساسية للمشروع بالوظائف التالية:

- أ- تقدير الوضع الحالي للشركة، ودراسة وسطي الإيرادات الشهرية، وتحليل سلوك الزبائن.
- ب- تصنيف الزبائن بحسب عدة محددات، منها: بياناتهم الشخصية (العمر، الجنس، الوظيفة، ...)، وسلوكهم تجاه الخدمات والعروض، وقيمة فاتورتهم الشهرية.
- ت- إيجاد العلاقات بين الأصناف السابقة، كالعلاقة بين جنس الزبون وسلوكه تجاه الخدمات والعروض (الإناث أكثر انجذاباً)، أو العلاقة بين عمر الزبون وقيمة الفاتورة.
- ث- إيجاد السعر الأنسب للحزم المراد طرحها بحيث تستقطب أكبر عدد من الزبائن، وتزيد من إيرادات الشركة، والمفاضلة بينهما على أساس سياسة الشركة.
- ج- إيجاد التوزيع الأفضل للخدمات ضمن الحزمة.
- ح- دراسة الإيرادات المتوقعة عند طرح أيّ من الحزم السابقة.
- خ- دراسة تأثير الحزم ببعضها للوصول إلى أفضل الأسعار للحزم، بحيث تحقق أفضل الإيرادات، وتجذب عدد أكبر من الزبائن.

٢-٣. مدخلات النظام BI-TS

يشكل مستودع البيانات (DW) الذي قمنا ببنائه الدّخل الأساسي للنظام BI-TS. الجدير بالذكر هنا أننا قمنا بتصميم وبناء مستودع البيانات (DW) قبل تطوير وبناء النظام BI-TS، بسبب وجود البيانات بشكلها المجرد في مصادر مختلفة، وعدم وجود مستودع مركزي للبيانات (Enterprise Data Warehouse). تضمنت هذه العملية جميع الأنشطة اللازمة لتجميع البيانات في قاعدة بيانات مرحليّة، ومكاملتها، وتنظيفها، وتصميم مستودع البيانات (DW)، وبنائها، وتحويل البيانات إليه من قاعدة البيانات المرحليّة، وسنقوم خلال هذا الفصل بتوضيح بنية كل منهما.

٢-٤. مخرجات النظام BI-TS

يشكل مخزن المعرفة والمعلومات (IKR) الخرج الأساسي للنظام BI-TS، حيث سيتم تخزين المعلومات والمعارف المكتشفة خلال كل دورة من مراحل العمل ضمنه، للاستفادة منها خلال الدورات اللاحقة للعمل. "يجمع مخزن المعرفة والمعلومات (IKR) كل من المعرفة الداخليّة والخارجيّة، والتي تخزن في قاعدة بيانات علائقية عموماً" (Turban and Aronson, 2001). سنوضح خلال هذا الفصل بنية هذا المخزن، وآليّة تخزين المعلومات والمعارف المكتشفة ضمنه.

٢-٥. البيئات البرمجية المستخدمة

قمنا باستخدام الأدوات والبيئات البرمجية التالية خلال مراحل تطوير وبناء النظام BI-TS:

- نظام إدارة قواعد البيانات (Microsoft Sql Server 2008): لبناء قاعدة البيانات المرحلية التي تم تجميع البيانات فيها، وبناء كل من مستودع البيانات (DW) ومخزن المعرفة والمعلومات.
- نظام توليد التقارير (SQL Server Reporting Services: SSRS): لتوليد التقارير والمخططات البيانية التي توضح نتائج عمل النظام.
- بيئة التطوير البرمجي (Microsoft Visual Studio): لبناء وتطوير بوابة تقوم بالوصول للبيانات المخزنة في قواعد البيانات، وتحليلها، وإظهار النتائج من خلال التقارير التي تم بناؤها على المخدم (SSRS).

٣. مراحل العمل

جرت العديد من الاجتماعات مع عدد من العاملين في شركة سيرتل في أقسام عدة، منها: خدمة الزبائن، والمحاسبة والفواتير، نظم إدارة المعلومات، أمن المعلومات، التسويق. الهدف الأساسي من وراء هذه الاجتماعات هو تحديد الوظائف الأساسية للنظام، ومدخلاته، والمخرجات المتوقعة في نهاية العمل. تم تحديد خمس دورات أساسية لعمل النظام BI-TS يتم في الدورة الأولى تحليل بيانات الزبائن واستخدمهم للشبكة الخليوي وتصنيفهم بناءً عليها، وفي الدورتين اللاحقتين يتم تحديد أنسب سعر لحزمة واحدة مراد طرحها، وما تتضمنه من خدمات، وتأثيرها على الإيرادات والزبائن، وفي الدورتين الأخيرتين يتم دراسة أثر الحزم ببعضها عند طرح أكثر من حزمة، وتأثيرها على الإيرادات والزبائن. سنعتمد على عدد من موظفي شركة الاتصالات سيرتل للعب بعض الأدوار المضمنة خلال دورة حياة النموذج (SCRUM-BI)، بحيث يقوم موظفي خدمة الزبائن وموظفي قسم التسويق بالأدوار التالية: محلل الأعمال، وخبير مجال العمل، والمدير الاستراتيجي، ويقوم موظفي المحاسبة والفواتير بدور محلل البيانات، ويقوم موظفي نظم إدارة المعلومات بدور مهندس المعرفة، وسنقوم بلعب بقية الأدوار وهي: خبير المنهجية، ومهندس البيانات، والمنقب في البيانات.

سنقوم فيما يلي بتوضيح مراحل عمل النموذج (SCRUM-BI) لبناء النظام BI-TS الذي تم توصيفه

سابقاً:

٣-١. فهم بيئة العمل

تشكل هذه العملية نقطة انطلاق النموذج (SCRUM-BI) وواحدة من أهم العمليات فيه. نتم خلال هذه المرحلة من دورة حياة النموذج بتقدير الوضع الحالي لبيئة العمل وفهم أهداف ومتطلبات المشروع من منظور العمل، بالشكل الذي سيساعد في تعريف أهداف وفرضيات المشروع لاحقاً. كان لموظفي قسم التسويق وقسم خدمة الزبائن الدور الهام في تحديد أبرز المشاكل والصعوبات التي من الممكن لنظامنا المساعدة في مواجهتها، بالإضافة لتحديد المتطلبات والأهداف المنشود منه في مجال التسويق والاهتمام بالزبائن.

شركة الاتصالات سيرتل هي إحدى الشركات المشغلة للهاتف المحمول (الخليوي) في سورية، استطاعت منذ تأسيسها في العام ٢٠٠٠ بناء سمعتها الجيدة من خلال مشاركتها في الاجتماعية والاقتصادية، سعياً لكسب رضى الزبائن وجذب اهتمامهم. لدى سيرتل عشرات مراكز الخدمة الموزعة في جميع أنحاء الجمهورية العربية السورية، وثلاث مراكز اتصالات أساسية موزعة في كل من دمشق، وحلب، واللاذقية. تميزت بنموها السريع؛ فهي تملك مئات الموظفين الذين يقدمون الخدمات لأكثر من ٧ مليون زبون. وتقدم خدمتي الجيل الثاني (2G) والجيل الثالث (3G) عبر محطاتها الراديوية المنتشرة في معظم أنحاء الجمهورية العربية السورية.

التركيز الأساسي لشركة سيرتل من وراء اعتماد تطبيقات ذكاء الأعمال (BI) هو الاهتمام بالزبائن، ومتطلباتهم، والمحافظة عليهم، وجذب زبائن آخرين، وذلك بالكشف عن فرص تنافسية ممكنة في سوق العمل. حيث تساعد هذه التطبيقات في تصنيف الزبائن، وتحديد تصاميم مناسبة للعروض والخدمات المقدمة.

تم تحديد مجموعة النقاط الأساسية الواجب على النظام BI-TS تحقيقها، من أهمها:

- الاهتمام ينصب على الخطوط لاحقة الدفع.
- تحليل الواقع الحالي للشركة (الإيرادات والزبائن).
- تصنيف الزبائن بحسب مجموعة من المحددات (الملف الشخصي وسلوكهم على شبكة الاتصالات).
- تقديم اقتراحات في مجال طرح العروض ومحتواها، ودراسة وتحليل أثرها على الزبائن والإيرادات عند طرح حزمة واحدة أو أكثر من حزمة معاً.

٣-٢. فهم البيانات

تشكل العملية الأساسية الثانية من دورة حياة النموذج (SCRUM-BI)، يتم خلالها الوصول لقواعد البيانات، وتحليلها، وتوصيفها، وضمان جودتها. كان لموظفي قسم المحاسبة والفواتير الدور الأساسي خلالها، كونهم المشغلون الأساسيون لقواعد بيانات الزبائن، وفواتيرهم، واستهلاكهم.

تُقسم البيانات في شركة سيرتل للاتصالات إلى الأنماط الثلاث التالية:

(١) بيانات تفصيل المكالمات: تمثل تفاصيل المكالمات التي يجريها الزبائن، وتشمل معلومات كافية لوصف الخواص الهامة لكل مكالمات، فكل تسجيل مكالمات ستضمن على الأقل الحقول التالية: رقم الهاتف الذي أجرى المكالمات، رقم الهاتف الوجهة، تاريخ المكالمات، مدتها، كلفتها، آلية تسعيرها.

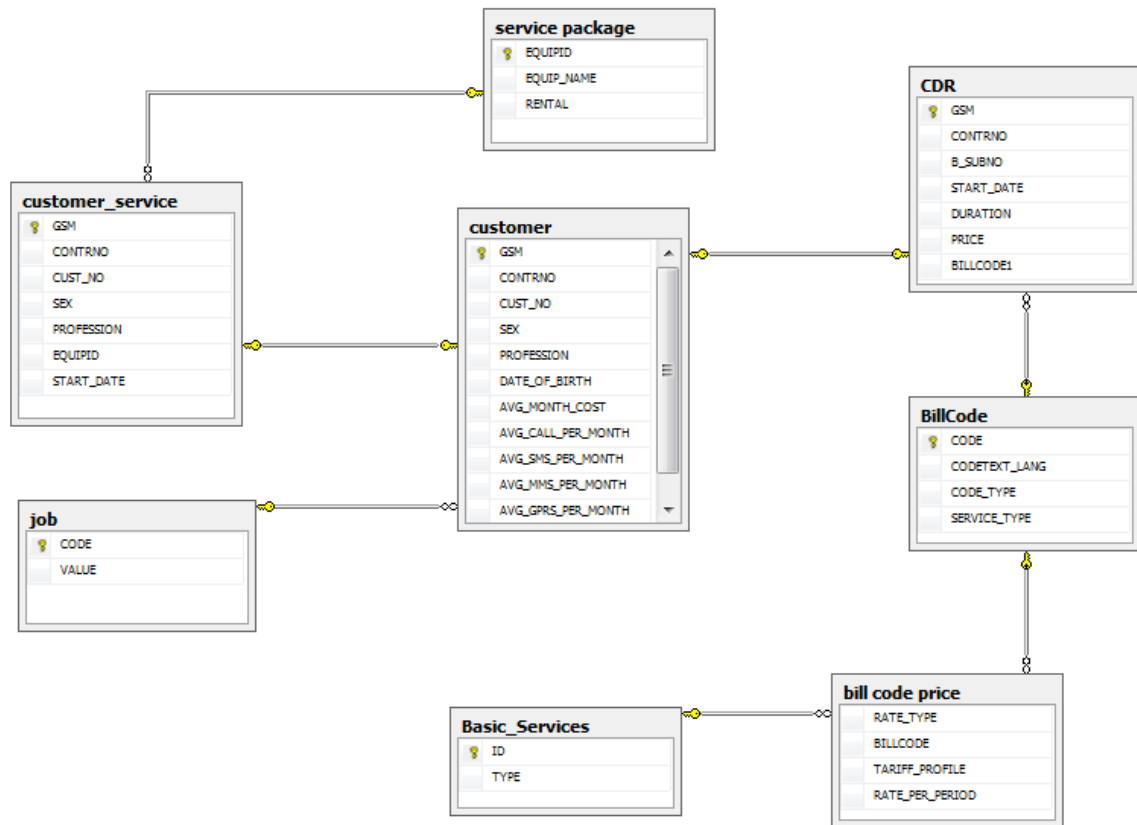
(٢) بيانات الزبائن: تختص بالبيانات الشخصية لزبائن الشركة، ويمكن الاستفادة منها في تطبيقات التنقيب عن البيانات، لتحليل البيانات الشخصية للزبائن، وتصنيفهم، وتحليل سلوكهم، وما يقدمه ذلك من دعم للعاملين في مجال التسويق أو خدمة الزبائن، للمحافظة على الزبائن الحاليين، وتقديم العروض المناسبة لجذب زبائن جدد.

(٣) بيانات شبكة الاتصالات: تتضمن هذه البيانات معلومات حول العتاد الصلب أو العتاد البرمجي المركب على تجهيزات شبكة الاتصالات، وبيانات بالرسائل المؤلدة منها، وأسباب هذه الرسالة.

حصلنا خلال هذه المرحلة على بيانات حوالي ٤٠٠٠ زبون -مستخدمي خطوط لاحقة الدّفع- وجميع عملياتهم على شبكة سيرتل الخليوية لمدة ستة أشهر (من تاريخ ٢٠١١/١١/١ وحتى تاريخ ٢٠١٢/٤/٣٠).

تم خلال هذه العملية تجميع البيانات التي تم الحصول عليها بشكل ملفات نصية في قاعدة بيانات مرحلية،

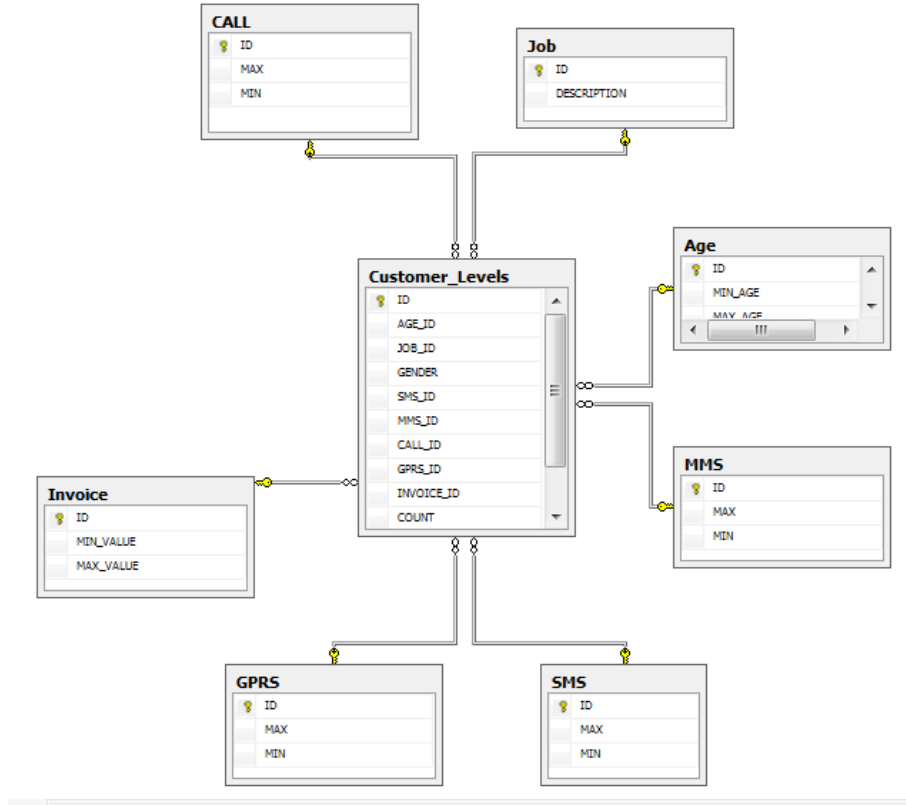
والموضحة بالشكل (٢٢) التالي:



الشكل (٢٢) مخطط علاقات الكيانات (ERD) لقاعدة البيانات المرحلية

يُعد الجدولين (Customer, CDR) أهم الجداول في قاعدة البيانات هذه، فيحوي الجدول Customer

بيانات الزبائن الشخصية لحوالي ٤٠٠٠ زبون، ويحوي الجدول CDR بيانات مكالماتهم على شبكة سيرتل، والتي تجاوزت ١٠ مليون تسجيلية (Record) ضمن قاعدة البيانات. قمنا بجميع عمليات المعالجة، والتنظيف، والمكاملة على هذه البيانات، قبل تحويلها لمستودع البيانات (DW) المستخدم في نظامنا والموضح بالشكل (٢٣)، تم تصميم مستودع البيانات (DW) بناءً لمتطلبات النظام، التي تقوم على تصنيف الزبائن على مجموعة من المحددات، وتتضمن محددات الملف الشخصي للزبون، ومحددات وسطي الاستخدام الشهري للخدمات.



الشكل (٢٣) مخطط علاقات الكيانات (ERD) لمستودع البيانات (DW) المستخدم في الحالة الدراسية

٣-٣. وضع الأهداف وتحديد الفرضيات

تُعتبر هذه العملية آخر عملية من عمليات التحضير للنموذج والتنقيب في البيانات، تمّ خلالها الاجتماع مع موظفي شركة سيرتل في مختلف الأقسام (خدمة الزبائن، والمحاسبة والفواتير، ونظام إدارة المعلومات، والتسويق، وغيرهم)، لتوصيف أهداف النظام BI-TS في كل دورة من دورات العمل، وتحديد الفرضيات والقيود في كل منها.

تهدف شركات الاتصالات من تطبيقات ذكاء الأعمال (BI) والتنقيب في البيانات لتوصيف زبائنهم، للاستفادة من هذه التوصيفات لاحقاً في عمليات التسويق والترويج التجاري، حيث يتم تقسيم الزبائن أصحاب السمات والخصائص المتشابهة في مجموعات متجانسة، مما يمنح الشركة القدرة على التفاعل بشكل أكبر مع الزبائن، وتحديد الجهود التسويقية وتوجيهها بالشكل المناسب.

وضعت خلال هذه المرحلة مجموعة من الأهداف والمتطلبات خلال دورات النظام BI-TS، من أبرزها:

- دراسة وتحليل الوضع الحالي للشركة، وتتضمن:

- ١) تصنيف الزبائن بحسب عدة محددات، منها: بياناتهم الشخصية (العمر، الجنس، الوظيفة، ...)، وسلوكهم تجاه الخدمات والعروض، وقيمة فاتورتهم الشهرية.
- ٢) إيجاد العلاقات بين الأصناف السابقة، كالعلاقة بين جنس الزبون وسلوكه تجاه الخدمات والعروض (الإناث أكثر إنجاباً)، أو العلاقة بين عمر الزبون وقيمة الفاتورة.
- ٣) دراسة وسطي الإيرادات الشهرية بحسب الأصناف التي تم تحديدها سابقاً.
- دراسة وتحليل حالة طرح حزمة من الخدمات، وتتضمن:
 - ١) دراسة وتحديد السعر الأنسب للحزمة المراد طرحها.
 - ٢) دراسة وتحديد حجم الإيرادات المتوقعة نتيجة طرحها.
 - ٣) دراسة وتحديد عدد الزبائن المتوقع استفادتهم من الحزمة.
 - ٤) تحليل أصناف الزبائن المستفيدين من الحزمة.
 - ٥) إيجاد التوزيع الأفضل للخدمات ضمن الحزمة.
 - ٦) توقع حجم الإيرادات عند طرح الحزمة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
 - ٧) توقع عدد الزبائن المستفيدين من الحزمة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
 - ٨) توقع الشهر الأفضل لطرح الحزمة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
- دراسة وتحليل حالة طرح أكثر من حزمة من الخدمات في وقت واحد، وتتضمن:
 - ١) دراسة تأثير الحزم ببعضها للوصول إلى أفضل الأسعار للحزم.
 - ٢) دراسة وتحديد السعر الأنسب لكل حزمة من الحزم المراد طرحها.
 - ٣) دراسة وتحديد حجم الإيرادات المتوقعة نتيجة طرح الحزم السابقة.
 - ٤) دراسة وتحديد عدد الزبائن المتوقع استفادتهم من كل حزمة من الحزم المراد طرحها.
 - ٥) تحليل أصناف الزبائن المستفيدين من كل حزمة من الحزم المراد طرحها.
 - ٦) إيجاد التوزيع الأفضل للخدمات ضمن كل حزمة من الحزم المراد طرحها.
 - ٧) توقع حجم الإيرادات عند طرح الحزم السابقة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
 - ٨) توقع عدد الزبائن المستفيدين من الحزم السابقة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
 - ٩) توقع الشهر الأفضل لطرح الحزم السابقة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها.
 - ١٠) دراسة الإيرادات المتوقعة عند طرح أيّ من الحزم السابقة.

جميع النتائج المتعلقة بالأهداف التي تم تحديدها في هذه المرحلة سيتم عرضها من خلال مجموعة كبيرة من التقارير، والمخططات البيانية، والمؤشرات، ولوحات التحكم، بالشكل الذي يمكن من عرض النتائج وإظهارها بشكل واضح ومفهوم، وسيتم تخزين جميع النتائج التي تم تحديدها كنتائج صحيحة، أو ضمن مجال التوقع في مخزن المعرفة والمعلومات (IKR).

٣-٤. اختيار البيانات

تمّ في هذه العملية اختيار البيانات المستهدفة، والتي سيتم تنفيذ خوارزميات التنقيب في البيانات عليها. إنّ البيانات المخزنة في مستودع البيانات (DW) قد تم تحميلها من عينة البيانات التي تم الحصول عليها من شركة سيرتل للاتصالات، وهذه العينة تم اختيارها بناءً لأهداف ووظائف النظام BI-TS، كما أنّ حجمها متوسط نسبياً -بحدود ١٠ مليون تسجيلة- مما يسمح بتحليلها ومذجتها كاملةً، لذلك اقتصرت عملية اختيار عينات البيانات على وظائف الإسقاط (Projection) والتجميع (Aggregation)، باستخدام مجموعة من تقنيات تخفيض الأبعاد (Dimensionality reduction) بحسب المحددات اللازمة في عملية تحليل البيانات والتنقيب فيها، كتجميع الزبائن بناءً لمتوسط فاتورتهم الشهرية عند تحليل السعر الأنسب للحزمة المراد طرحها.

٣-٥. النمذجة والتنقيب في البيانات

تشكل هذه العملية المكوّن الأساسي في دورة حياة النموذج (SCRUM-BI)، تم خلالها تصنيف الزبائن استخدام عدد من خوارزميات التنقيب والنمذجة كالتصنيف (Classification)، والعنقدة (Clustering)، وقواعد الارتباط (Association rules). كما تم تطبيق عدد من الأساليب والخوارزميات الإحصائية لتحليل العائدات بناءً للأصناف السابقة، وتوقع السعر الأنسب للحزم المراد طرحها، وتحديد توزيع الخدمات ضمن كل حزمة، وتحليل العائدات قبل طرحها وبعدها.

سنحاول فيما يلي تلخيص أبرز الحالات التي تم العمل عليها في النظام:

(١) دراسة وتحديد السعر الأنسب للحزمة المراد طرحها

بعد القيام بعمليات التصنيف والعنقدة للزبائن بحسب بياناتهم الشخصية وسلوكهم تجاه الخدمات، يمكننا الاستفادة من هذه التصنيفات في تحليل السعر الأنسب للحزمة المراد طرحها بحسب وسطي الفاتورة الشهرية للزبائن، ويمكن تلخيص خطوات العمل بالخطوات التالية:

أ- تصنيف الزبائن بحسب متوسط الفاتورة الشهري.

ب- حساب عدد الزبائن والإيرادات من أجل كل سعر للحزمة بحسب متوسط فاتورتهم الشهري.

- ت- إظهار مخططات بيانية توضح عدد الزبائن وحجم الإيرادات المتوقع من أجل كل سعر.
ث- اقتراح السعر الأفضل بحسب عدد الزبائن أو حجم الإيرادات المتوقع.

٢) إيجاد التوزيع الأفضل للخدمات ضمن الحزمة

بعد اختيار السعر الأنسب للحزمة المراد طرحها، قمنا بتحليل أصناف الزبائن الذين سيستفيدون من هذه الحزمة، لتحديد التوزيع الأفضل للخدمات ضمن هذه الحزمة، وذلك اعتماداً على تصنيفهم بحسب وسطي استخدام هؤلاء الزبائن للخدمات المدروسة في نظامنا (مكاملة صوتية، رسالة نصية، رسالة وسائط متعددة، تصفح الويب)، مرت هذه الحالة من النظام BI-TS بالخطوات التالية:

- أ- تصنيف الزبائن بحسب الوسطي الشهري لاستخدامهم للخدمات المدروسة.
ب- تحديد عدد من النماذج لتوزع الخدمات ضمن الخدمة (١٥ نموذج تغطي مختلف الحالات).
ت- حساب عدد الزبائن من أجل كل نموذج، وذلك بناءً لسعر الحزمة بالنسبة لوسطي فاتورته الشهرية من جهة، ونسبة توافق تفاصيل النموذج مع متطلبات الزبون ووسطي استخدامه للخدمات المدروسة.
ث- إظهار مخططات بيانية توضح عدد الزبائن وحجم الإيرادات من أجل كل نموذج.
ج- اقتراح التوزيع الأفضل للخدمات ضمن الحزمة بحسب عدد الزبائن أو حجم الإيرادات المتوقع.

٣) توقع حجم الإيرادات عند طرح الحزمة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها

تمّ في هذه الحالة من نظامنا دراسة الإيرادات المتوقعة بعد طرح الحزمة التي تمّ تحديد سعرها وتوزع الخدمات ضمنها. تمر هذه الحالة بالخطوات التالية:

- أ- حساب الإيرادات الشهرية المتوقعة من طرح الحزمة.
ب- مقارنة هذه الإيرادات مع الإيرادات الأساسية للشركة.
ت- تقديم مقترحات بزيادة سعر خدمة ما في حال تجاوز محتويات الحزمة.

٤) توقع الشهر الأفضل لطرح الحزمة بعد تحديد السعر الأنسب لها والتوزيع الأفضل للخدمات ضمنها

قمنا خلال هذه الحالة من نظامنا بتحديد الشهر الأنسب لطرح حزمة ما من أجل جميع نماذج الحزم المدروسة، حيث يتبع النظام BI-TS الخطوات التالية:

- أ- تحديد أصناف الزبائن الأكثر تأثراً بالحزمة المراد طرحها، وذلك بناءً لسعر الحزمة وتوزع الخدمات ضمنها.
ب- مقارنة بينة الخدمات ضمن الحزمة مع السلوك الشهري للزبائن ضمن الأصناف المحددة سابقاً.
ت- حساب احتمال استفادة هؤلاء الزبائن من الحزمة المراد طرحها في كل شهر.
ث- إظهار مخططات بيانية لعدد الزبائن المتوقع استفادتهم من الحزمة في كل شهر.

ج- اقتراح الشهر الأنسب لطرح الحزمة بحسب عدد الزبائن أو حجم الإيرادات المتوقع.

٥) دراسة وتحليل حالة طرح أكثر من حزمة من الخدمات في وقت واحد

نحاول في هذه الحالة تحليل أثر طرح أكثر من حزمة في آنٍ واحد على الزبائن والإيرادات، ودراسة تأثير هذه الحزم ببعضها. تمر هذه الحالة بالخطوات التالية:

أ- تحديد عدد الحزم المراد طرحها.

ب- تحديد مستويات الفاتورة المطلوب طرح الحزمة فيها.

ت- تحديد قائمة الأسعار الممكنة في كل مستوى.

ث- اختيار السعر الأنسب لكل حزمة بحسب عدد الزبائن أو حجم الإيرادات المتوقع.

ج- تحليل نموذج توزيع الخدمات ضمن كل حزمة بحسب عدد الزبائن أو حجم الإيرادات المتوقع.

٦) دراسة التغييرات على سلوك الزبائن عند إعادة تسعير خدمة ما

تم في هذه الحالة حساب التغييرات على استخدام الزبائن للخدمات عند إعادة تسعير هذه الخدمات، وذلك بناءً لتصنيفات هؤلاء الزبائن بالنسبة للوسطي الشهري لاستخدامهم لهذه الخدمات. يمر النظام BI-TS في هذه الحالة بالخطوات التالية:

أ- تصنيف الزبائن بحسب الوسطي الشهري لاستخدامهم للخدمة ما (رسائل نصية مثلاً).

ب- حساب التغييرات على سلوك الزبائن نتيجة تغيير سعر الخدمة.

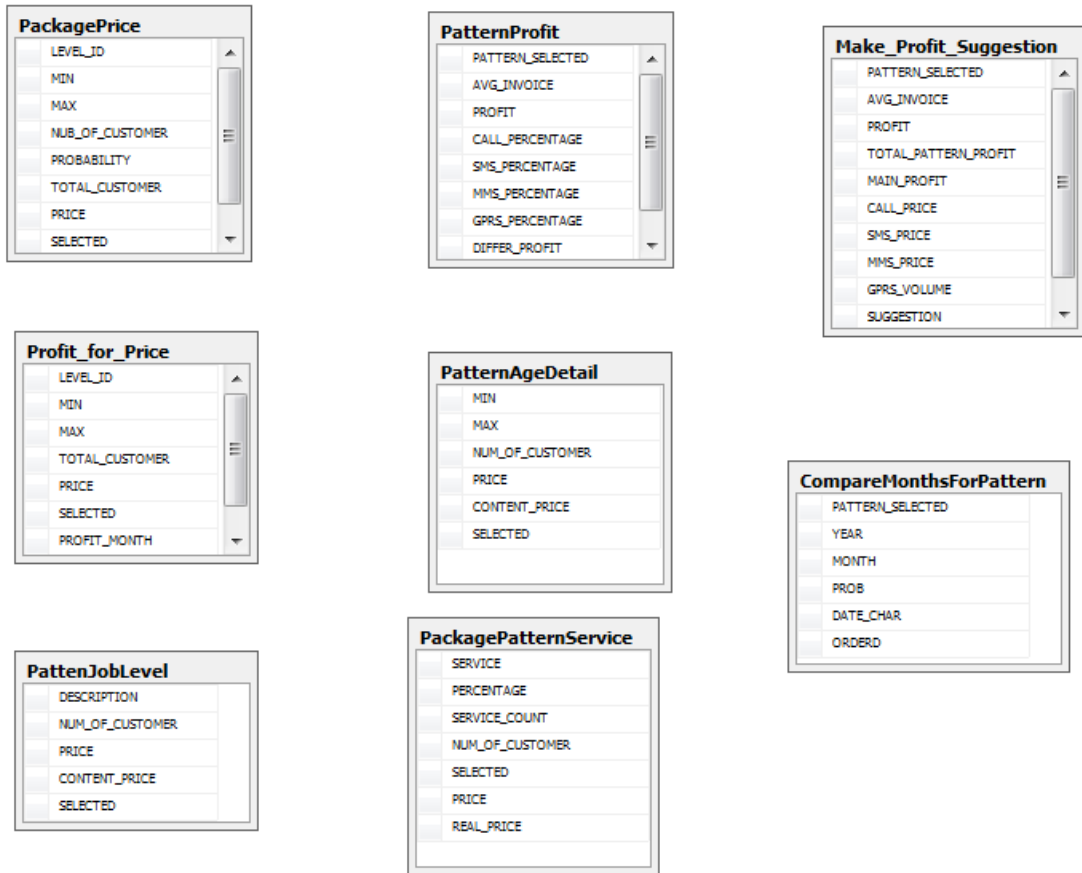
ت- إظهار مخططات بيانية توضح الفروقات بين الاستخدام المتوقع للخدمة نتيجة السعر الجديد، والوسطي الشهري لاستخدام الخدمة عند السعر الحالي.

٣-٦. التقييم

قمنا خلال هذه العملية من دورة حياة النموذج بالتحقق من صحة أساليب النمذجة وخوارزميات التنقيب في البيانات المستخدمة، والنتائج التي تم الحصول عليها قبل نشرها في العملية اللاحقة، وضمان تحقيق أهداف النظام BI-TS، والتي تم وضعها وتحديدها في مرحلة سابقة. تطلبت هذه العملية تعاوناً من قبل موظفي شركة سيرتل في مختلف الأقسام، وذلك لضمان أنّ النتائج ضمن مجال التوقع، ومتوافقة مع حالات سابقة، كما تمّ الحصول على بعض ردود الأفعال المفيدة لتصحيح عمل النظام أو تعديله، بالشكل الذي يمكننا من تحقيق الأهداف المرجوة من هذا النظام. نذكر هنا على سبيل المثال لا الحصر إحدى الملاحظات المهمة وهي: أظهرت بعض النتائج لإحدى نماذج الحزم وجود خسارة بعد طرحها، وتم سؤال موظفي التسويق وموظفي نظم إدارة المعلومات حولها، الذين أجابوا بوجود هكذا حالات، ولذلك

يقومون برفع أسعار الخدمات بعد تجاوز الزبون للحجم المضمن ضمن العرض أو الحزمة، لذلك قمنا بالعودة لعملية النمذجة والتنقيب بالبيانات، وقمنا بدراسة أثر إعادة تسعير خدمة ما على سلوك الزبائن تجاه هذه الخدمة.

تخزن جميع النتائج التي تم الحصول عليها في مخزن المعرفة والمعلومات (IKR)، وذلك بغض النظر عن نشرها في العملية اللاحقة أو عدمه، وذلك للاستفادة منها في دورات النظام BI-TS الذي نقوم ببنائه وتطويره أو في مشاريع أخرى للتنقيب في البيانات في شركة سيرتل. ويوضح الشكل (٢٤) بنية مخزن المعلومات والمعرفة (IKR) وهو عبارة عن قاعدة بيانات علائقية كما ذكرنا سابقاً.



الشكل (٢٤) مخطط علاقات الكيانات (ERD) لمخزن المعلومات والمعرفة (IKR) المستخدم في النظام BI-TS

٣-٧. النشر

تنتهي كل دورة من دورات حياة النموذج (SCRUM-BI) بعملية نشر النتائج، قمنا في هذه العملية بتطوير بوابة للنظام BI-TS، والموضحة بالشكل (٢٥)، تمكن من الوصول لأجزاء النظام BI-TS المختلفة والانتقال بينها بسهولة، وتسمح بعرض النتائج بطريقة واضحة ومفهومة من خلال مجموعة واسعة من المخططات البيانية، والمؤشرات، ولوحات التحكم.

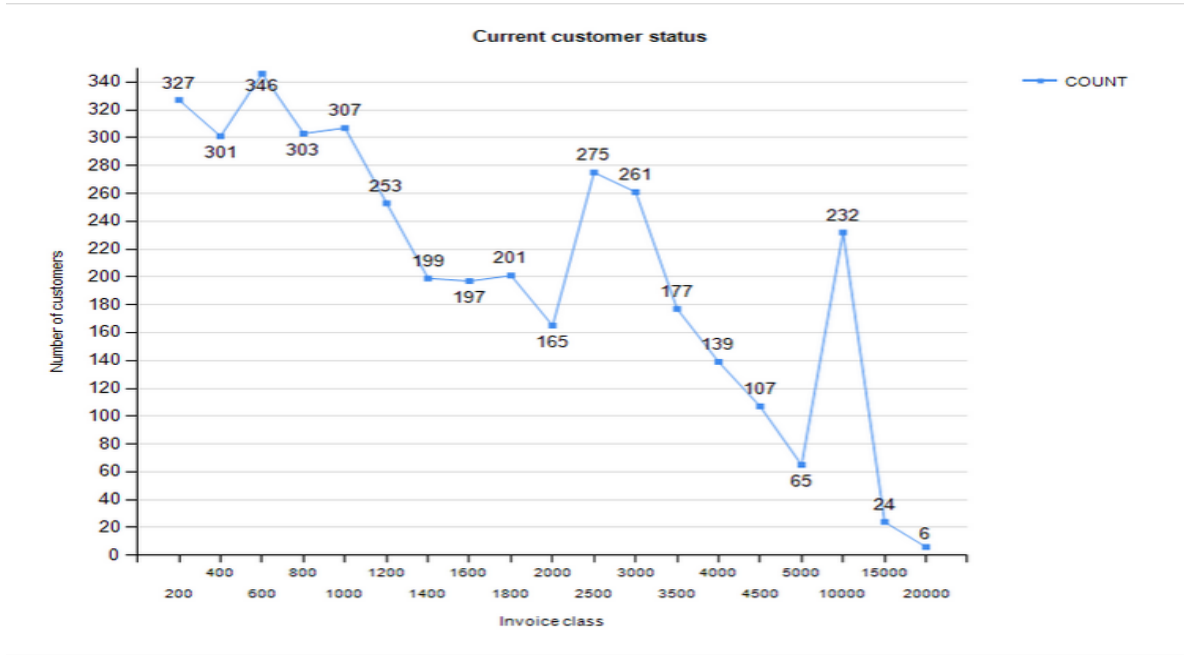


الشكل (٢٥) بوابة النظام (BI-TS)

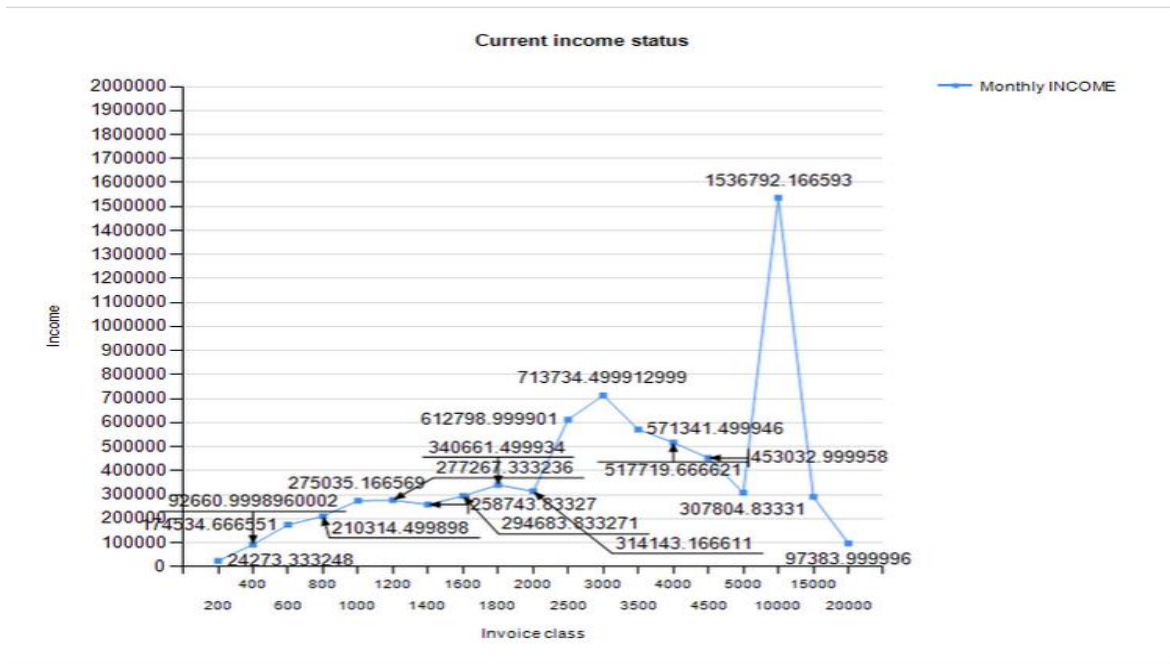
يقدم هذا النظام المساعدة والدعم للعاملين في شركة سيرتل بشكل عام، وموظفي قسم التسويق بشكل خاص من خلال نقطتين أساسيتين هما: (I) تقديم صورة شاملة عن واقع الشركة الحالي، وزبائنها، وتفاعلهم على شبكة الاتصالات الخاصة بها. (II) المساعدة في العمليات التسويقية والترويجية، بالشكل الذي يساهم في جذب الزبائن، والمحافظة عليهم، وزيادة إيرادات الشركة. سنقوم فيما يلي بعرض لمحة عن اهم مخرجات النظام BI-TS ضمن هاتين النقطتين:

(I) تقديم صورة شاملة عن واقع الشركة الحالي:

يظهر النظام BI-TS مجموعة من المخططات البيانية التي تظهر وضع الشركة بالنسبة لعدد الزبائن كما هو موضح بالشكل (٢٦)، وبالنسبة للإيرادات الشهرية كما هو موضح بالشكل (٢٧).



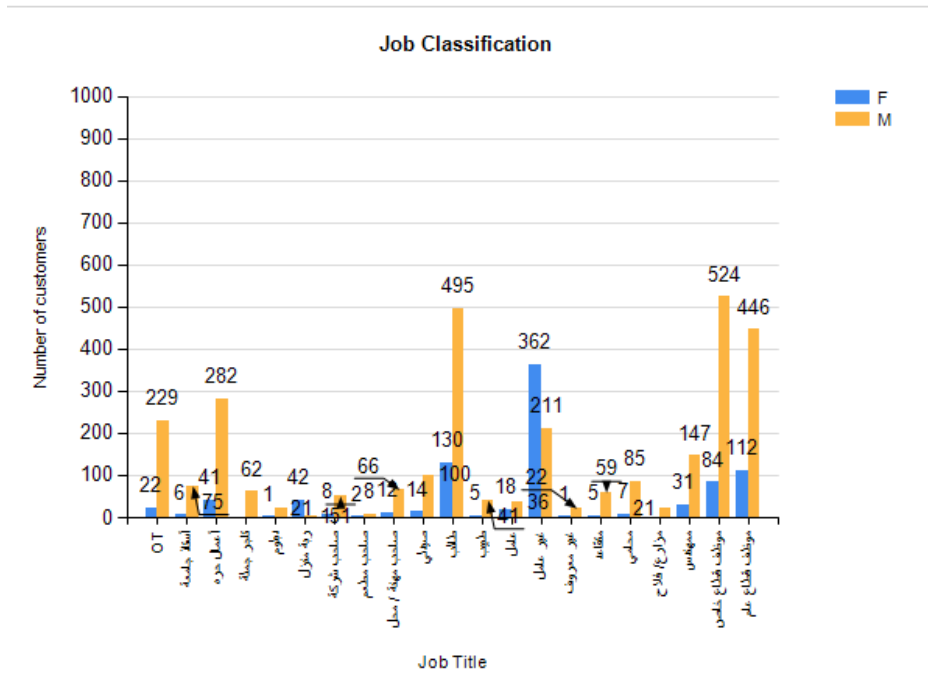
الشكل (٢٦) مخطط بياني لعدد الزبائن بالنسبة لمتوسط الفاتورة الشهرية



الشكل (٢٧) مخطط بياني لحجم الإيرادات الشهرية

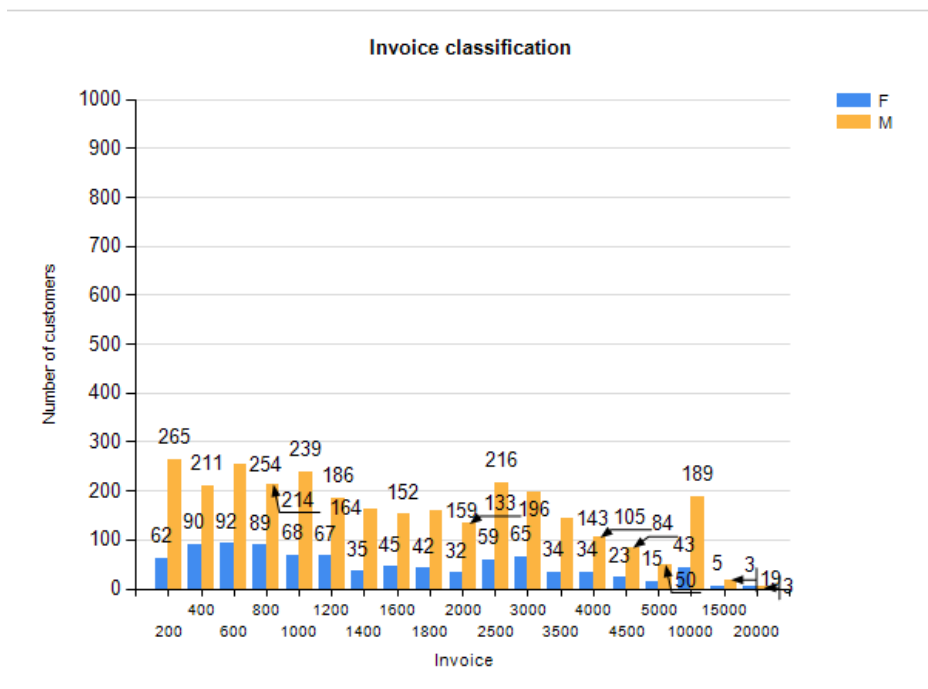
كما يقوم النظام BI-TS بتصنيف الزبائن بالنسبة لبياناتهم الشخصية أو سلوكهم على شبكة الاتصالات، وإظهار التقارير والمخططات البيانية المناسبة اللازمة، لتصنيف الزبائن بالنسبة للمهنة في الشكل (٢٨)، وتصنيفهم بالنسبة لمتوسط فاتورتهم الشهرية في الشكل (٢٩).

This report show customer classification by Job



الشكل (٢٨) مخطط بياني لتصنيف الزبائن بالنسبة للمهنة والجنس

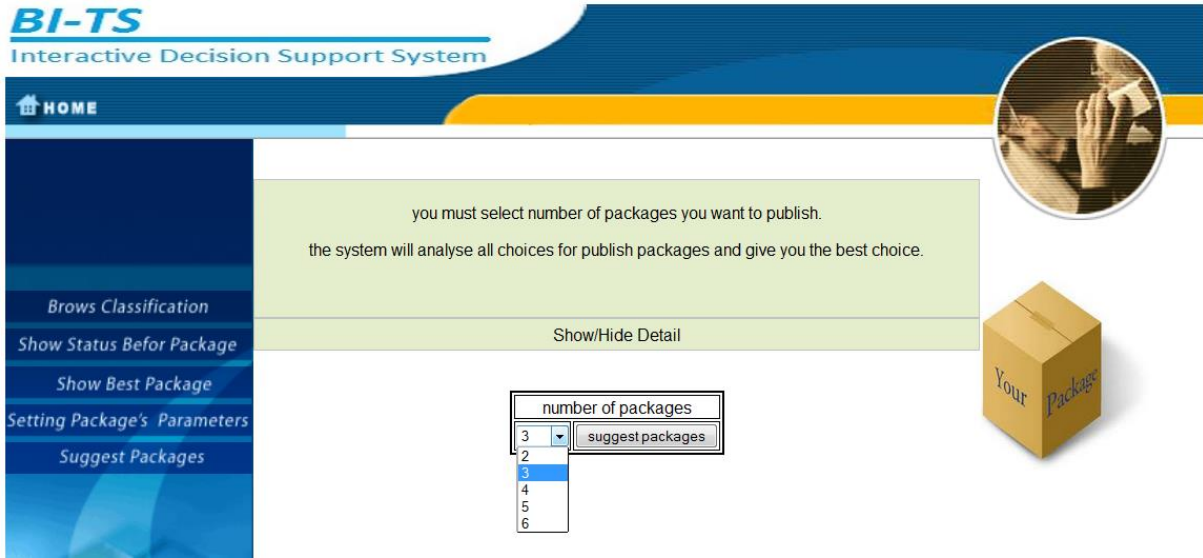
This report show show customer classification by Invoice



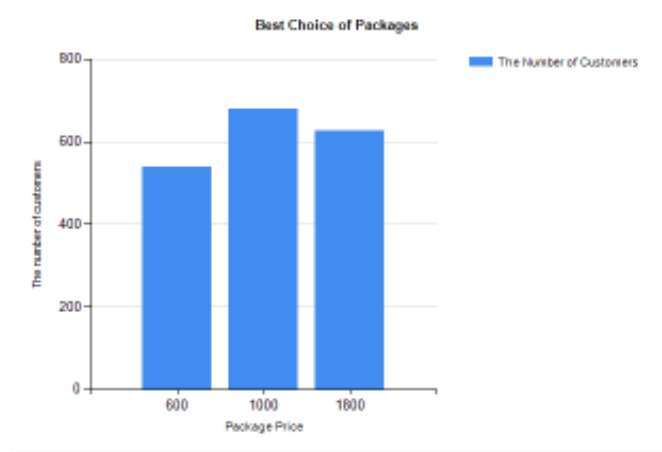
الشكل (٢٩) مخطط بياني لتصنيف الزبائن بالنسبة لمتوسط الفاتورة الشهرية

(II) المساعدة في العمليات التسويقية والترويجية:

يقدم النظام BI-TS مجموعة من التقارير والمخططات البيانية، والتي تساعد في عملية طرح العروض وحزم الخدمات، فيمكن تحديد عدد الحزم المراد طرحها كما في الشكل (٣٠)، ليقوم النظام BI-TS بعرض أفضل الأسعار لكل حزمة بحسب عدد الزبائن المتوقع اشراكهم في كل منها، بالإضافة لجدول يظهر تفاصيل الخيارات الأخرى، كما في الشكل (٣١).



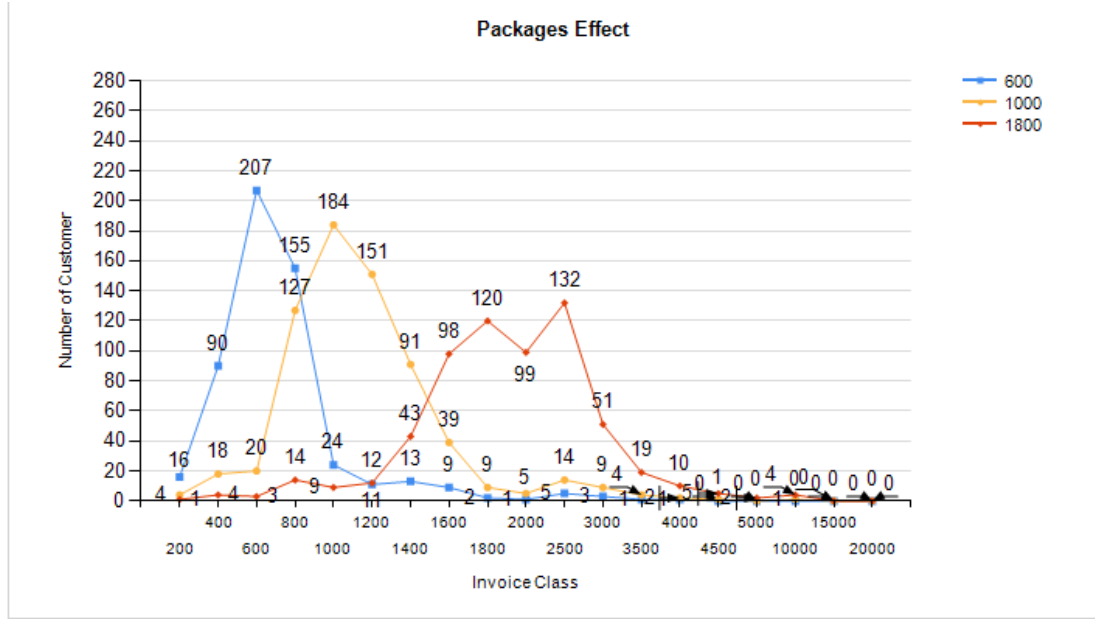
الشكل (٣٠) تحديد عدد الحزم المراد طرحها



STATUE Number	Package Price	Number of Customers
1	Total	1843
2	Total	1830
	600	557
	1200	652
	1700	611
3	Total	1823

الشكل (٣١) أفضل الأسعار لكل حزمة بحسب عدد الزبائن المتوقع اشراكهم في كل منها

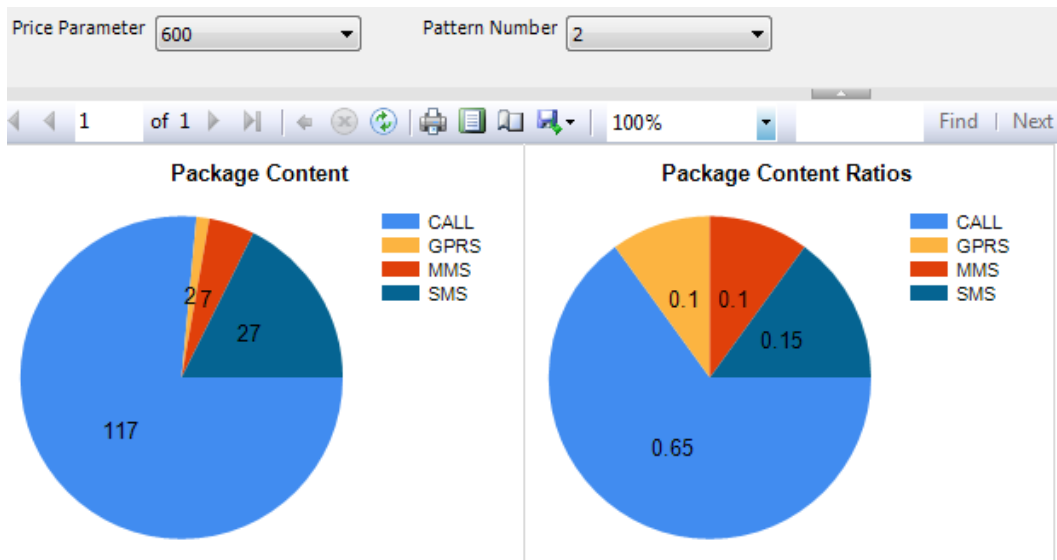
كما يظهر النظام BI-TS مخططاً بيانياً لتأثيرات الحزم المقترحة ببعضها البعض بالنسبة للزبائن المتوقع اشتراكهم في كل منها، كما يوضح الشكل (٣٢).



PACKAGE PRICE	LEVEL ID	Number of Customers
600	Total	538
1000	Total	679

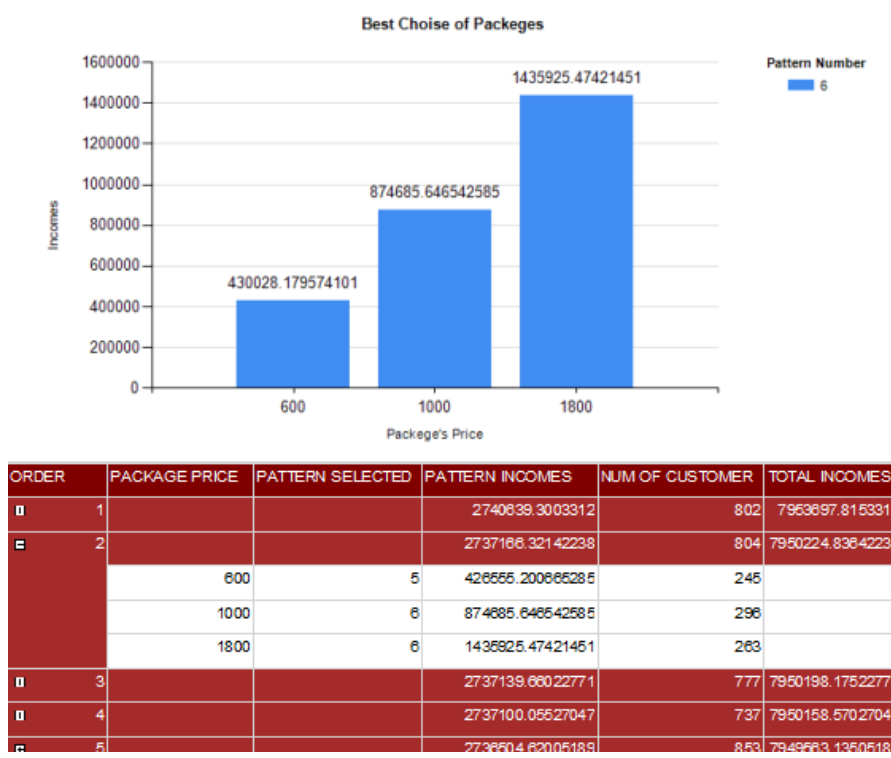
الشكل (٣٢) مخطط بياني يظهر تأثيرات الحزم المقترحة ببعضها البعض بالنسبة للزبائن المتوقع اشتراكهم في كل منها

يمكن للنظام BI-TS تحديد نموذج توزيع الخدمات ضمن كل حزمة من الحزم المقترحة، كما في الشكل (٣٣).

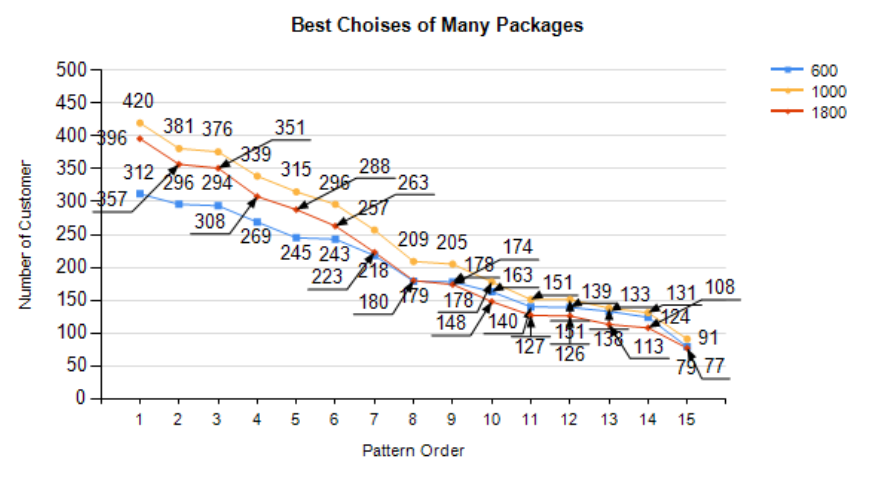


الشكل (٣٣) نموذج توزيع الخدمات ضمن كل حزمة من الحزم المقترحة

يتيح النظام BI-TS بالإضافة لذلك تحليل إيرادات النماذج السابقة وعدد الزبائن المتوقع اشتراكهم بها من خلال التقارير الموضحة بالشكل (٣٤). والشكل (٣٥).



الشكل (٣٤) إيرادات الحزم المقترحة



الشكل (٣٥) مخطط بياني لعدد الزبائن المتوقع اشتراكهم في كل حزمة من الحزم المقترحة

الجدير بالذكر هنا أنّ البوابة (BI-TS) تقدم عدد كبير من التقارير والمخططات البيانية التي لا يمكن عرضها

جميعا في هذا البحث.

٤. خاتمة

قدمنا في هذا الفصل لمحة عن نظام ذكاء الأعمال (BI) الذي قمنا ببنائه (النظام BI-TS) خلال الحالة الدراسية، محاولين توضيح التطبيق العملي للنموذج (SCRUM-BI) المقدم في هذا العمل. حاولنا خلال هذا الفصل توضيح الإطار العام للنظام BI-TS، ومراحل تطويره التي مرّ بها خلال دورة حياة النموذج SCRUM-BI، بحيث أجبنا على بعض فرضيات البحث المعرّفة سابقاً من هذا العمل، ومبرزين كيف يساهم النموذج SCRUM-BI في بناء تطبيقات ذكاء الأعمال (BI)، والتي يمكن أن تساعد المدراء ومتخذي القرار في تقييم الوضع الحالي للشركة، ودعم قراراتهم الاستراتيجية في مجالات التسويقية والترويجية لها.

الفصل السادس

الخاتمة والآفاق المستقبلية

قدمنا في هذا العمل نموذجاً جديداً لاكتشاف المعرفة في البيانات (KDD) سميناه النموذج (SCRUM-BI). النموذج (SCRUM-BI) يقوم على الدمج والمزاوجة بين نماذج اكتشاف المعرفة تقليدية التهج والمنهجية الرشيقة لتطوير البرمجيات المنهجية سكروم، بالشكل الذي يساعد في بناء وتطوير تطبيقات ذكاء الأعمال (BI) في بيئات عمل متطورة بشكل متسارع، ومتغيرة باستمرار على مستوى المتطلبات والأولويات. يسهم النموذج (SCRUM-BI) في تحسين وتعزيز عملية الحصول على المعرفة ومشاركتها، بالشكل الذي يساعد في دعم عملية اتخاذ القرارات الاستراتيجية. إنَّ النموذج (SCRUM-BI) يستفيد من نقاط القوة في النماذج السابقة، ويتغلب على نقاط الضعف فيها. تم مراعاة جوانب عدة في بيئة النموذج (SCRUM-BI)، وهي: محازن البيانات التي تشكل دخل وخرج النموذج (SCRUM-BI)، وطبيعة العمليات خلال كل دورة من دورات حياة النموذج (SCRUM-BI)، وطبيعة الموارد البشرية المضنة في النموذج (SCRUM-BI)، وطبيعة المسارات المتبعة خلال دورة حياة النموذج (SCRUM-BI)، والتكيف مع المتغيرات في بيئات العمل والمتطلبات، والبعد الاستراتيجي للنموذج (SCRUM-BI) ودوره في دعم عمليات اتخاذ القرارات الاستراتيجية، والبعد المعرفي للنموذج (SCRUM-BI) ودوره في تحصيل المعرفة، وتنظيمها، وتخزينها، ومشاركتها، وإظهارها.

تم اختبار النموذج (SCRUM-BI) باستخدام حالة دراسية في شركة سيرتل، قمنا خلالها ببناء نظام لذكاء الأعمال (BI) سميناه (النظام BI-TS). يقدم النظام (BI-TS) الدعم والمساعدة لمتخذي القرار في شركة سيرتل لتحليل وضع الشركة الحالي، ويدعم عملية اتخاذ القرارات الاستراتيجية في مجالات التسويقية والترويجية لها.

قمنا خلال هذا العمل بتوصيف بنية مخزن المعرفة والمعلومات (IKR) المستخدم في النموذج (SCRUM-) (BI)، ولكنّه لا يزال يواجه قصوراً في طريقة بناء هذا المخزن بشكل فعّال، مما يجعل هذه النقطة إحدى الآفاق المستقبلية التي يمكن العمل عليها. من النقاط الأخرى -التي يمكن العمل عليها مستقبلاً- الاستفادة من الأنطولوجية في تحسين عمليات اكتشاف المعرفة في البيانات (KDD)، فأنطولوجيات مجال العمل يمكن أن تساعد في تحسين عمليات فهم المشكلة وتحليلها، كما يمكن أن تسهم في بناء تطبيقات التنقيب في البيانات لاكتشاف المعرفة من البيانات بشكل آليّ أو شبه آليّ.

المراجع

- [1] G. P. Shapiro, C. Matheus and W. Frawley, "Knowledge Discovery in Databases - An Overview," *AI Magazine*, pp. 57-70, September 1992.
- [2] U. Fayyad, G. P. Shapiro and P. Smyth, *Advances in knowledge discovery and data mining*, Menlo Park, CA: American Association for Artificial Intelligence, 1996.
- [3] G. P. Shapiro, "Analytics and data mining: The key to successful CRM," in *Knowledge Discovery and Data Mining Conference*, Boston, MA, 2000.
- [4] E. Knapp and C. Havens, "Easing into knowledge management," *Strategy & Leadership*, pp. 4-9, March 1999.
- [5] S. M. Weiss, S. J. Buckley, S. Kapoor and S. Damgaard, "Knowledge-based data mining," in *KDD'03 Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington, DC, USA, August 24 - 27, 2003.
- [6] W. McKnight, "Ask the CRM expert," 2002. [Online]. Available: http://expertanswercenter.techtarget.com/eac/knowledgebaseAnswer/0,295199,sid63_gci974430,00.html.
- [7] M. Alnoukari and A. El Sheikh, *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications*, USA: IGI Global, 2012, p. 370.
- [8] V. Dhar and R. Stein, *Seven methods for transforming corporate data into business intelligence*, Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1997.
- [9] Capgemini, "Agile Business Intelligence How to make it happen," White Paper, 2014.
- [10] M. Hofmann and B. Tierney, "An enhanced data mining life cycle," in *Computational Intelligence and Data Mining, 2009. CIDM '09. IEEE Symposium on*, Nashville, TN, 2 April 2009.

- [11] N. McMurchy, "Toolkit Tactical Guideline: Five Success Factors for Effective BI Initiatives," Gartner, Inc., 2008.
- [12] KDnuggets, "Data Mining ROI," May 2008. [Online]. Available: <http://www.kdnuggets.com/polls/2008/roi-data-mining.htm>.
- [13] E. R. Keith, "Agile Software Development Processes - A Different Approach to Software Design," 2002.
- [14] H. P. Luhn, "A business intelligence system," *IBM Journal of Research and Development*, pp. 314-319, October 1958.
- [15] J. Mitja, L. Nada and U. Tanja, "Managing business intelligence in a virtual enterprise: A case study and knowledge management lessons learned," *Journal of Intelligent & Fuzzy Systems*, vol. 14, no. 3, pp. 121-136, 2003.
- [16] S. Negash, "Business Intelligence," *Communications of the Association for Information Systems*, vol. 13, no. 1, pp. 177-195, 15 February 2004.
- [17] J. Hightower and M. Shariat, "Conceptualizing Business Intelligence Architecture," *Marketing Management Journal*, vol. 17, no. 2, pp. 40 - 46, 1 September 2007.
- [18] P. Marren, "The father of business intelligence," *Journal of Business Strategy*, vol. 25, no. 6, November 2004.
- [19] N. Kerdprasop and K. Kerdprasop, "Moving Data Mining Tools toward a Business Intelligence System," *International Journal of Intelligent Technology*, vol. 2, no. 2, pp. 99-104, 2007.
- [20] B. Azvine, Z. Cui and D. D. Nauck, "Towards real-time business intelligence," *BT Technology Journal*, vol. 23, no. 3, pp. 214-225, July 2005.
- [21] H. Kemper and H. Baars, "Management Support with Structured and Unstructured Data—An Integrated Business Intelligence Framework," *Information Systems Management*, vol. 25, no. 2, pp. 132-148, 2008.

- [22] E. Turban, J. E. Aronson, T. P. Liang and R. Sharda, *Systems, Decision Support and Business Intelligence*, vol. 8, Upper Saddle River, NJ: Prentice-Hall, Inc., 2006.
- [23] R. Kimball, L. Reeves, W. Thornthwaite, M. Ross and W. Thornwaite, *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses with CD Rom*, 1st ed., New York, NY, USA: John Wiley & Sons, Inc., 1998, p. 800.
- [24] L. Mei and C. A. Thole, "Data analysis for parallel car-crash simulation results and model optimization," *Simulation Modelling Practice and Theory*, vol. 16, no. 3, p. 329–337, March 2008.
- [25] Agile Alliance, "Manifesto for Agile Software Development," 2001. [Online]. Available: <https://www.agilealliance.org/agile101/the-agile-manifesto/>.
- [26] K. Schwaber and M. Beedle, *Agile Software Development with Scrum*, 1st ed., Upper Saddle River, NJ: Prentice Hall PTR, 2001, p. 158.
- [27] J. A. Highsmith, *Adaptive software development: a collaborative approach to managing complex systems*, New York, NY, USA: Dorset House Publishing Co., Inc., 2000, p. 358.
- [28] K. Beck, *Extreme programming explained: embrace change*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1999, p. 190.
- [29] A. Cockburn, *Agile software development*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2002, p. 278.
- [30] J. Stapleton, *Dsdm: The Method in Practice*, Boston, MA, USA: Addison-Wesley Longman Publishing Co, Inc., 1997, p. 192.
- [31] S. R. Palmer and J. M. Felsing, *A Practical Guide to Feature-driven Development*, Prentice Hall PTR, 2002, p. 271.
- [32] S. Ambler, *Agile Modeling: Effective Practices for eXtreme Programming and the Unified Process*, New York, NY, USA: John Wiley & Sons, Inc., 2002, p. 400.

- [33] K. Schwaber and J. Sutherland, "The Definitive Guide to Scrum: The Rules of the Game," White Paper, The Scrum Guide™, 2011.
- [34] Z. Stojanovic, A. Dahanayake and H. Sol, "Modeling and Architectural Design in Agile Development Methodologies," in *Proceedings of the 8th CAISE/IFIP8.1 International Workshop on Evaluation Methods in System Analysis and Design EMMSAD '03*, Velden, Austria, 2003.
- [35] P. Abrahamsson, J. Warsta, M. T. Siponen and J. Ronkainen, "New directions on agile methods: a comparative analysis," in *Proceedings of the 25th International Conference on Software Engineering, ICSE '03*, Washington, DC, USA, 2003.
- [36] M. A. Feldens, M. R. Leal De, A. Pavan and L. S. Brazil, "Towards a Methodology for the Discovery of Useful Knowledge Combining Data Mining, Data Warehousing and Visualization," in *Proceedings of the conferencia Latinoamericana de informatica quito ecuador*, 1998.
- [37] A. G. BÜCHNER, M. D. MULVENNA, S. S. ANAND and J. G. HUGHES, "An internet-enabled knowledge discovery process," in *Proceedings of the 9th international database conference*, Hong Kong, 1999.
- [38] P. Gottgroy, "An Ontology Driven Knowledge Discovery Framework for Dynamic," New Zealand, 2010.
- [39] L. A. Kurgan and P. Musilek, "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review*, vol. 21, no. 1, pp. 1-24, 1 March 2006.
- [40] SEMMA, "Data mining and the case for sampling," White Paper, SAS Institute Inc., 1997.
- [41] SPSS Modeler, "CRISP-DM Guide," White Paper, IBM Corporation., 2008.
- [42] K. J. Cios, W. Pedrycz, R. W. Swiniarski and L. Kurgan, *Data Mining: A Knowledge Discovery Approach*, 1st ed., New York, USA: Springer US, 2007, p. 606.

- [43] K. Collier, B. Carey, E. Grusy, C. Marjaniemi and D. Sautter, "A perspective on data mining," Northern Arizona University, Arizona, United States, 1998.
- [44] A. Azevedo and M. F. Santos, "KDD, semma and CRISP-DM: A parallel overview," in *Proceedings of IADIS European Conference on Data Mining 2008*, Amsterdam, The Netherlands, July 2008.
- [45] CRISP-DM, "Step-by-step data mining guide," White Paper, CRISP-DM Consortium, 2000.
- [46] K. Rennolls and A. AL-Shawabkeh, "Formal structures for data mining, knowledge discovery and communication in a knowledge management environment," *Intelligent Data Analysis*, vol. 12, no. 2, pp. 147-163, 16 April 2008.
- [47] M. Hofmann and B. Tierney, "Development Phases of a Generic Data Mining Life Cycle (DMLC)," in *Proceedings of the International Conference on Software Engineering Theory and Practice*, Orlando, Florida, USA, July 2007.
- [48] H. P. Brasza, "Expert Systems Technology," White Paper, Churchill Systems, Inc., 2000.
- [49] S. C. Yoon, S. Makki, L. J. Henschen and E. Park, "Using domain knowledge in knowledge discovery," in *CIKM '99 Proceedings of the eighth international conference on Information and knowledge management*, New York, NY, USA, 1 November 1999.
- [50] J. Phillips and B. G. Buchanan, "Ontology-guided knowledge discovery in databases," in *K-CAP '01 Proceedings of the 1st international conference on Knowledge capture*, New York, NY, USA, 2001.
- [51] P. Gottgroy, N. Kasabov and S. Macdonell, "An ontology engineering approach for Knowledge Discovery from data in evolving domains," in *Proceedings of the 4th International Conference on Data Mining (Data Mining IV)*, Rio de Janeiro, Brazil, 2003.
- [52] V. Svátek, J. Rauch and M. Flek, "Ontology-based explanation of discovered associations in the domain of social reality," in *2nd*

ECML/PKDD Workshop on Knowledge Discovery and Ontologies, Porto, Portugal, 2005.

- [53] V. Svátek, J. Rauch and M. Ralbovský, "Ontology-Enhanced Association Mining," in *Semantics, Web and Mining*, Berlin Heidelberg, Springer, 2006, pp. 163-179.
- [54] Y. T. Kuo, A. Lonie, L. Sonenberg and K. Paizis, "Domain ontology driven data mining: a medical case study," in *DDDM '07 Proceedings of the 2007 international workshop on Domain driven data mining*, New York, NY, USA, 12 August 2007.
- [55] P. Gottgroy, "Ontology driven Knowledge discovery process: a proposal to integrate ontology engineering and KDD," in *PACIS 2007 Proceedings of the 11th Pacific-Asia Conference on Information Systems*, Auckland, New Zealand, July 2007.
- [56] A. Linden, "Free Methodology and Process Model for Data Mining Released," Gartner, Inc., 2000.
- [57] S. W. Lee and L. Kerschberg, "A methodology and life cycle model for data mining and knowledge discovery in precision agriculture," in *Systems, Man, and Cybernetics, 1998. 1998 IEEE International Conference on*, San Diego, CA, Oct 1998.

ملحق ١

نماذج اكتشاف المعرفة الأخرى والنماذج الأكثر استخداماً

كما أشرنا سابقاً في هذا البحث هناك العديد من نماذج اكتشاف المعرفة في البيانات (KDD) بدايةً من نموذج (Fayyad et al., 1996)، وقد قمنا في دراستنا المرجعية باختيار أهم هذه النماذج بناءً على مجالات تطبيقها واستخدامها في المجالين العلمي والصناعي.

نورد في هذا الملحق عدداً من نماذج اكتشاف المعرفة التي تم استبعادها إما لتأثيرها الضعيف من جهة، أو لاعتمادها على إحدى المنهجيات التي تم مناقشتها في دراستنا المرجعية. الجدول التالي يلخص بعض سمات هذه النماذج: الاسم والمرجع والنهج المتبع وعدد الخطوات وخطوات عمل كل نموذج.

ملخص سمات نماذج اكتشاف المعرفة المستبعدة من الدراسة المرجعية (التقري، ٢٠١٢)

KDP Model	Ref.	KDP Approach	No of Steps	KDP Model Steps
Adriaans & Zantinge	(Adriaans and Zantinge 1996)	Traditional KDP Approach	6	Data Selection, Cleaning, Enrichment, Coding, DM, Reporting.
Berry & Linoff	(Berry and Gordon 1997)	Traditional KDP Approach	4	Identifying the Problem, Analyzing the Problem, Taking Action, Measuring the Outcome.
Feldens et al.	Feldens et al. (1998)	Traditional KDP Approach	3	Pre-Processing, Data Mining, Post-Processing.
Cabena et al.	(Cabena et al. 1998)	Traditional KDP Approach	5	Business Objectives Determination, Data Preparation, DM, Domain Knowledge Elicitation, Assimilation of Knowledge.
Edelstein	(Edelstein 1998)	Traditional KDP Approach	5	Identifying the Problem, Preparing the Data, Building the Model, Using the Model, Monitoring the Model.
Anand & Buchner	Buchner et al. (1999)	Web-based KDP Approach	8	Human Resource Identification, Problem Specification, Data Prospecting, Domain Knowledge Elicitation, Methodology Identification, Data Preprocessing, Pattern Discovery, Knowledge Post-processing.
Reinartz	(Reinartz 1999)	Traditional KDP Approach	7	Business Understanding, Data Understanding, Data Preparation, Data Exploration, Data Mining, Evaluation, Deployment.
Kopanakis & Theodoulidis	Kopanakis and Theodoulidis (1999)	Traditional KDP Approach	6	Scrub, verify and summarize data, Selection of the training data sample, Model derivation algorithm, Verify and evaluate, Selection of most interesting models, Model usage + population shift monitoring and incremental learning.






Han & Cercone	(Han and Cercone 2000)	Traditional KDP Approach	5	Original Data Visualization, Data Reduction, Data Preprocess, Pattern Discovery, Pattern Visualization.
Cios et al.	(Cios et al. 2000)	Traditional KDP Approach	6	Understanding the Problem Domain, Understanding the Data, Preparation of the Data, DM, Evaluation of the Discovered Knowledge, Using the Discovered Knowledge.
Han & Kamber	(Han and Kamber 2001)	Traditional KDP Approach	9	Learning the Application domain, Creating a Target Data Set, Data Cleaning and Preprocessing, Data Reduction and Transformation, Choosing Functions of DM, Choosing the Mining Algorithm(s), DM, Pattern Evaluation and Knowledge Presentation, Use of Discovered Knowledge.
Klosgen & Zytkow	(Klosgen and Zytkow 2002)	Traditional KDP Approach	7	Definition and Analysis of Business Problems, Understanding and Preparation of Data, Setup of the Search for Knowledge, Search for Knowledge, Knowledge Refinement, Application of Knowledge in Solving the Business Problems, Deployment and Practical, Evaluation of the Solutions.
Haglin et al.	(Haglin et al. 2005)	Traditional KDP Approach	7	Goal Identification, Target Data Creation, Data Preprocessing, Data Transformation, DM, Evaluation and Interpretation, Take Action steps.
Pabarskaite & Raudys	Pabarskaite and Raudys (2007)	Web-based KDP Approach	9	Data collection, Data cleaning, User identification, Session identification, Feature selection, Data transformation, Data combination, Mining the data, Result visualization.
Li & Ruan	(Li and Ruan 2007)	Traditional KDP Approach	6	Data Collection, Selection, Preprocessing, Transformation, Data Mining, Interpretation/Evaluation.
KDCK	Rennolls and AL-Shawabkeh (2008)	Ontology-based KDP Approach	5	Data Collection and Processing, Data Understanding, DM/Modeling, Knowledge Understanding, Business Understanding.

كما نورد في هذا الملحق احصائية أجريت من قبل <http://www.kdnuggets.com> **KDnuggets**

تبين النماذج الأكثر استخداماً. الإحصائية أجريت في العام ٢٠٠٧ وتضمنت أكثر من ١٥٠ جواب، وتم إعادة التصويت في العام ٢٠١٤ وتضمنت أكثر من ٢٠٠ جواب وكانت النتائج على الشكل التالي:

الإحصائية المقدمة من موقع Kdnuggets لأكثر نماذج اكتشاف المعرفة استخداماً

What main methodology are you using for your analytics, data mining, or data science projects? [200 votes total]	
14 poll 14 poll	
CRISP-DM (86)	43% 42%
My own (55)	27.5% 19%
SEMMA (17)	8.5% 13%
Other, not domain-specific (16)	8% 4%
KDD Process (15)	7.5% 7.3%

My organizations' (7)	 3.5%  5.3%
A domain-specific methodology (4)	 2%  4.7%
None (0)	0%  4.7%

ملحق ٢

المنهجية الرشيقة سكروم

SCRUM

تعتبر المنهجية سكروم (SCRUM) إحدى أهم المنهجيات الرشيقة، وأكثرها انتشاراً واعتماداً من قبل المؤسسات والمنظمات حول العالم كما أسلفنا في هذا البحث. نقدم في هذا الملحق لمحة عن المنهجية سكروم؛ مؤضحين تطور مفهوم سكروم، وبينه المنهجية.

١. تطوّر مفهوم سكروم (SCRUM)

نشأ مصطلح سكروم (SCRUM) في لعبة الرّكي فهو يشير لاستئناف المباراة بعد توقف طفيف، ليتم استخدامه لاحقاً في ألعاب الكمبيوتر للإشارة لمتابعة اللعبة بعد توقفها لسبب ما. وعرف (Takeuchi and Nonaka, 1986) سكروم أولاً في مجال تطوير المنتجات بأنها "استراتيجية تطوير المنتجات بشكل كلي ومرن، حيث يعمل فريق التطوير كوحدة متكاملة للوصول إلى هدف مشترك بدلاً من نهج المتابعة التقليدية". ليصف الباحثون لاحقاً نهجاً جديداً لتطوير المنتجات التجارية الذي من شأنه زيادة السرعة والمرونة، وذلك استناداً إلى حالات دراسية من شركات التصنيع في مجال صناعات السيارات وآلات التصوير والطباعة، وسمّوه النهج الشمولي أو لعبة الرّكي، فالعملية برمتها يقوم بها فريق واحد لتحقيق هدف واحد عبر مراحل متداخلة متعددة، حيث "يحاول الفريق أن يقطعوا المسافة كوحدة واحدة مع تمرير الكرة ذهاباً وإياباً". فيما استخدم (Schwaber, 1990) طرق تطوير البرمجيات المتقدمة والتي سيصبح اسمها سكروم في شركته "OOPSLA" لاحقاً. ليكون (Sutherland, Scumniotales and McKenna, 1993) أوّل من يقدم نهجاً متكاملًا في تطوير البرمجيات المتقدمة تحت اسم المنهجية سكروم، ويعتمده في شركته "Easel Corp". ليقدّم أخيراً (Sutherland and Schwaber, 1995) مقالة علمية تصف المنهجية سكروم وهندستها.

ليتم استخدام هذه المنهجية لاحقاً في كبرى شركات قطاع التكنولوجيا وتطوير البرمجيات: غوغل "Google"، مايكروسوفت "Microsoft"، فيليبس "Philips"، ياهو "Yahoo"، نوكيا "Nokia"، سيمنز "Siemens"، وغيرها.

٢. بنية المنهجية سكروم

تقوم المنهجية سكروم على فرق عمل صغيرة منظمة ذاتياً، بحيث يتم تجزئة الأعمال إلى أجزاء صغيرة يتم إتمامها خلال مجموعة من الحلقات المتتالية والمسماة أشواط، والتي سنقوم بتوضيح بنيتها لاحقاً. فقد "وضعت بنية المنهجية سكروم بحيث تراعي مبادئ الشفافية والفحص والمتابعة والتكيف مع المتغيرات" (Sutherland and Schwaber, 2013)، وتتألف هذه البنية من فرق سكروم من جهة، والأدوار، والقواعد، والمفاهيم، والأحداث المرتبطة بهم من جهة أخرى، كما يظهر الشكل التالي.

مكونات بنية المنهجية الرشيقية سكروم "SCRUM"

الأدوار (Roles)	المفاهيم (Artifacts)	الأحداث (Events)
<ul style="list-style-type: none"> مالك المنتج (Product Owner). فريق التطوير (Development Team). مدير السكروم (Scrum Master). 	<ul style="list-style-type: none"> إضافة المنتج (Increment). جعبة المنتج (Product Backlog). جعبة الشوط (Sprint Backlog). 	<ul style="list-style-type: none"> الشوط (Sprint). تخطيط الشوط (Sprint Planning). السكروم اليومي (Daily Scrum). استعراض الشوط (Sprint Review). الاسترجاع (Retrospective).

٢-١. الأدوار (Roles)

صممت فرق العمل في المنهجية سكروم لتحسين الإنتاجية والمرونة والابداع، فتميز بالتنظيم الذاتي بحيث تختار الأسلوب الأفضل لتحقيق أهدافها وإنجاز العمل بدلاً من الحاجة لإدارتها من قبل فريق خارجي، كما تتميز هذه الفرق بشموليتها الوظيفية فتملك جميع القدرات اللازمة لأداء عملها دون الاستعانة بأي مساعدة خارجية.

يتألف فريق عمل سكروم من:

١-١-٢. مالك المنتج (Product Owner)

هو شخص وحيد يعتبر مسؤولاً عن زيادة فعالية فريق التطوير وزيادة عملهم، ويعتبر الشخص الوحيد المسؤول عن إدارة جعبة المنتج، وما يتضمنه من عمليات ترتيب وتوضيح لعناصر جعبة المنتج، بالشكل الذي يضمن أنه واضح وشفاف بالنسبة للجميع، بحيث يُظهر العمل الذي يجب أن يقوم به فريق التطوير في كل مرحلة ويُمكنهم من فهمه مما يساهم بتحقيق أهداف العمل بشكل أفضل. فلا يسمح لأي شخص آخر التدخل في عمل فريق التطوير، كما لا يسمح لفريق التطوير بالعمل على أي شيء آخر ممكن أن يُفرضه عليه.

٢-١-٢. فريق التطوير (Development Team)

يتكوّن فريق التطوير من مجموعة من الاختصاصيين الذين يقومون بجميع الأعمال اللازمة لإطلاق إضافة المنتج في نهاية كل شوط، حيث يتم تشكيل الفريق واختياره بالشكل الذي يمكنهم من تنظيم وإدارة عملهم بأنفسهم، مما يخلق نوعاً من التآزر والتعاون فيما بينهم، فيحسن كفاءة الفريق وفعاليتته.

يتميز فريق التطوير بمجموعة من الخصائص والمميزات:

- الفريق منظم ذاتياً: فلا يمكن لأي أحد خارجي أن يتدخل في عملهم.
- الفريق متعدد المهارات والخبرات: فيملكون جميع الخبرات والقدرات اللازمة لإنهاء ونشر إضافة المنتج.
- عدده صغيره نسبياً (لا يتجاوز ٩ أعضاء) ليبقى رشيقاً وبعيداً عن حجم التنسيق الهائل الذي تحتاجه الفرق كبيرة الحجم، وكبير نسبياً (لا يقل عن ٣ أعضاء) لينجز عمل هام خلال الشوط، مما يقلل عدد الدورات ويزيد الإنتاجية.

٣-١-٢. مدير سكروم (Scrum Master)

هو شخص وحيد أيضاً يعتبر مسؤولاً عن تطبيق مفاهيم المنهجية سكروم، وأن مبادئ المنهجية ومفاهيمها وقواعدها مفهومة من قبل الجميع ومطبقة بالشكل السليم في المشروع. كما يساعد الأشخاص خارج المشروع بفهم كيفية التواصل والتفاعل مع فريق التطوير.

٢-٢. الأحداث (Events)

تُعرّف المنهجية سكروم مجموعة من الأحداث، بهدف تنظيم العمليات خلال العمل والتقليل من الاجتماعات غير المعرفة ضمنها، وتميز جميع الأحداث في المنهجية بأنها محدّدة الوقت، بحيث يكون لكل حدث مدة قصوى، بمجرد بدء الشوط لا يمكن اختصار هذه المدة أو إطالتها. بينما الأحداث الأخرى يمكن أن تنتهي بمجرد انتهاء العمل ويتحقق

الغرض من الحدث. وبخلاف الشوط نفسه -الذي يحتوي كل الأحداث الأخرى - كل حدث في المنهجية سكروم هو فرصة لتفحص شيء ما أو التكييف معه، فقد صممت هذه الأحداث لضمان والتكييف الشفافية والتدقيق والفشل في تطبيقها أي منها سيقبل من شفافية العمل وفرص التدقيق.

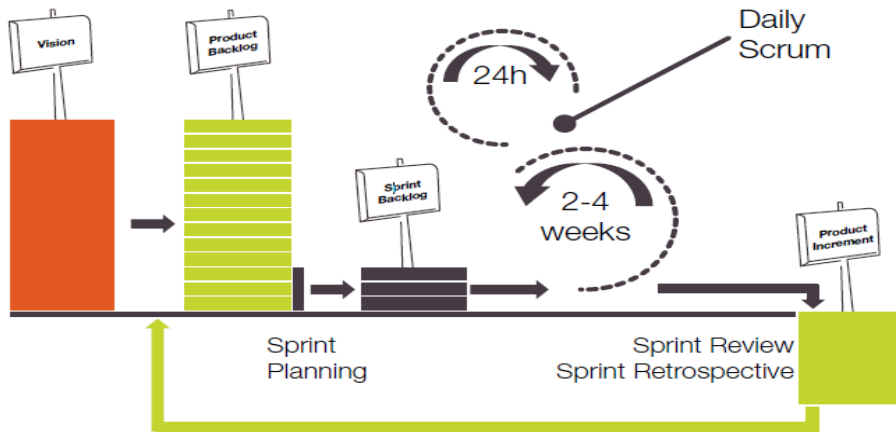
٢-٢-١. الشوط (Sprint)

يُعتبر الشوط قلب وأساس المنهجية سكروم، وهو محدد الزمن من أسبوع وحتى شهر، يتم خلاله بناء إضافة منتج تام وقابل للاستخدام والنشر. فيتكوّن الشوط من مجموعة من الأحداث التالية: تخطيط الشوط، وسكروم اليومي، والعمل التراكمي، واستعراض الشوط، واسترجاع الشوط. ولا يمكن إجراء أيّة تغييرات خلال الشوط، والتي من شأنها التأثير على هدفه وجودته.

يبدأ كل شوط جديد مباشرة بعد الانتهاء من الشوط السابق، فيمكن اعتبار كل شوط عبارة عن مشروع بمدة لا تتجاوز الشهر، حيث إنّ الشوط - كما المشروع - يهدف إلى إنجاز شيء ما؛ فيوجد تعريف واضح لما سيتمّ بناؤه، بالإضافة لوجود تصميم وخطة مرنة لقيادة عملية البناء هذه، وتنظيم المنتج الذي تمّ بناؤه خلال هذا الشوط.

يوضح الشكل التالي بنية الشوط وأهم الأحداث التي يتضمنها:

شوط سكروم "SCRUM Sprint"



٢-٢-٢. تخطيط الشوط (Sprint Planning)

في هذا الحدث يتم التخطيط لما سيتم القيام به خلال الشوط بالتعاون بين جميع أعضاء فريق سكروم الداخلي، بحيث نحصل في النهاية على توصيف دقيق لإضافة المنتج التي سنقوم بتسليمها في نهاية الشوط، والعمل اللازم القيام به لتحقيق ذلك.

يمتدّ هذا الحدث غالباً لثمانى ساعات كحد أقصى، وهنا تأتي مسؤولية مدير سكروم بالتعاون مع فريق سكروم للحفاظ على هذا الحد الزمني.

٢-٢-٣. سكروم اليوميّ (Daily Scrum)

من الواضح من خلال تسمية هذا الحدث بأنّه حدث يومي محدد الزمن بحدود خمس عشر دقيقة، يقوم خلالها فريق التطوير بمزامنة جميع الأنشطة التي يقومون بها والتخطيط للعمل المراد القيام به خلال اليوم التالي.

٢-٢-٤. استعراض الشوط (Sprint Review)

يتمّ استعراض الشوط عند نهايته، للتحقق من إضافة المنتج التي تمّ بناؤها وملاءمتها لجمعة المنتج ككل. يمتد هذا الحدث لمدة زمنية لا تتجاوز الأربع ساعات، يتعاون خلالها فريق سكروم وجميع الجهات المعنية لتقييم ما تمّ القيام به خلال الشوط، وذلك بحسب المهام التالية:

- مالك المنتج: يوضح ما تمّ إنجازه من جمعة المنتج وما لم يتمّ إنجازه، ويوضح تاريخ الانتهاء الكليّ للمنتج بناءً لما تمّ إنجازه حالياً.
- فريق التطوير: يشرح كيف سارت الأمور وما المشاكل التي واجهها وكيف تعامل معها من جهة، ويوضح الأعمال الخاصة بجمعة المنتج التي تمّ إنجازه، ويحيبوا على جميع الأسئلة حولها من جهة أخرى.

وفي النهاية يتعاون الفريق ككل لتحديد ما سيتم العمل عليه في الأشواط التالية لاحقاً، والتغيرات الطارئة على السوق حالياً، والأشياء الأفضل للقيام بها، بحيث يتمّ استعراض الإمكانيات الحالية والميزانية والجدول الزمنيّ، مما يوفر مساهمة قيّمة لتحديد دخل الشوط اللاحق.

٢-٢-٥. الاسترجاع (Retrospective)

يعتبر هذا الحدث فرصة لفريق سكروم لتفقد الوضع الحاليّ ووضع خطة مستقبلية لتحسينات في الأشواط اللاحقة، فيأتي هذا الحدث بعد حدث استعراض الشوط الحاليّ وقبل تخطيط الشوط التالي، ويمتدّ لمدة زمنية لا تتجاوز الثلاث ساعات يتمّ خلاله:

- تدقيق الشوط الحاليّ من كافة النواحي: الأشخاص المشاركون بالعمل، والعلاقات بينهم، والعمليات المتبعة، والأدوات المستخدمة.
- تحديد وترتيب العناصر التي تمّ إنجازها والتحسينات الممكنة عليها.
- الاستفادة من الخبرات المتراكمة في الشوط الحاليّ لوضع خطة لتحسين أداء فريق سكروم في الأشواط اللاحقة.

٢-٣. المفاهيم (Artifacts)

تعرف المنهجية سكروم مجموعة من المفاهيم لضمان الشفافية وخلق آليات أفضل للتدقيق والتكيف، حيث يتشارك فريق سكروم هذه المفاهيم وتعريفها الدقيق.

٢-٣-١. جعبة المنتج (Product Backlog).

تتكون جعبة المنتج من قائمة مرتبة لكل ما يحتاجه فريق سكروم لتوصيف المنتج، فهي المصدر الوحيد للمتطلبات الواجب على المنتج تحقيقها؛ فتحتوي جعبة المنتج قائمة بجميع المزايا والوظائف والمتطلبات للمنتج الحالي، والتحسينات المطلوب إدخالها على المنتج في الإصدارات اللاحقة. ولكل بند من هذه البنود مجموعة من السمات هي: التوصيف، والترتيب، والقيمة، والتقدير.

يعتبر مالك المنتج المسؤول الوحيد عن جعبة المنتج بما في ذلك: المحتوى والصلاحية والترتيب، فيما يعتبر فريق التطوير مسؤولاً عن جميع التقديرات في هذه الجعبة، وقد يساعد مالك المنتج بفهم هذه التقديرات والقيام ببعض المبادلات والتفاضلات، ليتم وضع التقدير النهائي بالتعاون بين فريق سكروم ككل.

٢-٣-٢. جعبة الشوط (Sprint Backlog).

يمكن تعريف جعبة الشوط بأنها مجموعة جزئية من جعبة المنتج لما سيتم تحقيقه في هذا الشوط، بالإضافة لخطة الحصول على إضافة المنتج وتحقيق هدف الشوط. فهي توقع من قبل فريق التطوير للوظائف التي سيتم إنجازها في إضافة المنتج المتوقعة نهاية الشوط، والعمل اللازم لإنجازها وتحقيق هدف الشوط.

فريق التطوير هو الوحيد مالك صلاحية التعديل على جعبة الشوط خلاله؛ فهو يقوم بإزالة ما هو غير ضروري من عناصرها، ويضيف أية أعمال جديدة عليها، ويعدّل تقديرات العمل المتبقي.

٢-٣-٣. إضافة المنتج (Product Increment).

يشتمل إضافة المنتج على جميع عناصر جعبة المنتج التي تم إنجازها خلال الشوط الحالي، ففي نهاية الشوط يجب أن تكون الإضافة الجديدة منجزة وقابلة للاستخدام سواء تقرر نشرها أو لم يتم نشرها.