علم البيانات الضخمة وتعلم الذكاء الإصطناعى

Big Data Science and Deep Learning

أ.د. معتز خورشيد جامعة القاهرة	أ.د. محمد العزب المنتدى العربى للذكاء الإصطناعى

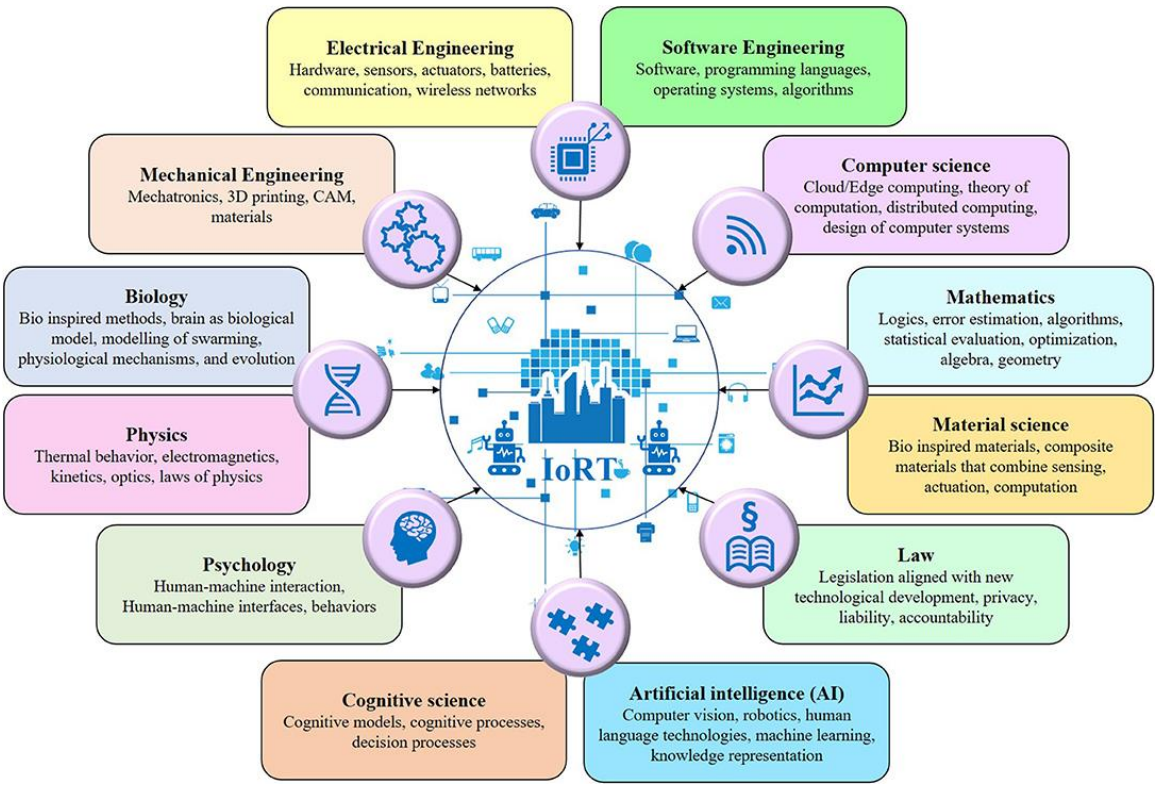
الملخص:

يمر العالم بسرعة نحو التحول الرقمي ويدخل إلى عالم جديد مع العديد من التطورات غير المسبوقة. وتؤثر هذه الاتجاهات الجديدة على المجتمع والوظائف المستقبلية، وبالتالي وظائف الخريجين الجدد. ويكمن في قلب هذا التحول الرقمي علم البيانات الضخمة؛ ذلك التخصص الذي يكسب البيانات الضخمة معنى. ومع وجود العديد من التحديات الرقمية الناشئة التي تتزايد معدلات نموها أمام أعيننا بسرعة فائقة ، تناقش هذه المقالة وجهات النظر حول فرص كليات المعلومات ، والمقترحات في مجال تعليم علم البيانات. ويمكن القول بأن كليات المعلومات يجب أن تمكن طلابها من خلال تخصصات "حوسبة المعلومات"، التي نعرّفها على أنها القدرة على حل المشكلات وإنشاء القيم والمعلومات والمعرفة باستخدام الأدوات في المجالات التطبيقية المختلفة. وباعتبارها مناهج متخصصة لدعم تخصصات حوسبة المعلومات في تعليم علم البيانات، فنقترح المحاور الثلاثة القائمة على المستخدم، والأداة، والتطبيق. ستعمل على التمييز بين تعليم علم البيانات في كليات المعلومات، وتعليم علوم الحاسب أو كليات إدارة الأعمال. ونقدم إطار تعليم علم البيانات متعدد الطبقات Data Science Education Framework (DSEF)، ذو وحدات بناء، تتضمن الركائز الثلاث لعلم البيانات (الأشخاص، والتقنية، والبيانات)، والتفكير الحاسوبي، والنماذج القائمة على البيانات، ودورات حياة علم البيانات. وبالتالي ، يجب تصميم مقررات علم البيانات استناداً إلى هذا الإطار باستخدام المناهج القائمة على المستخدم، والقائمة على الأدوات والقائمة على التطبيق. وسوف يساعد هذا الإطار طلابنا على التفكير في مشكلات علم البيانات من منظور الصورة الكبيرة، وتعزيز مهارات حل المشكلات المناسبة جنبًا إلى جنب مع المنظورات الواسعة لدورات حياة علم البيانات. ونأمل أن يساعد إطار تعليم علم البيانات متعدد الطبقات الذي تمت مناقشته في هذه المقالة الزملاء في كليات المعلومات في تصميمهم لمناهج علم البيانات الجديدة.

الكلمات المفتاحية: البيانات الضخمة، علم البيانات، حوسبة المعلومات، الثورة الصناعية الرابعة، مدارس المعلومات، التفكير الحاسوبي، نموذج مستند إلى البيانات ، دورة حياة علم البيانات.

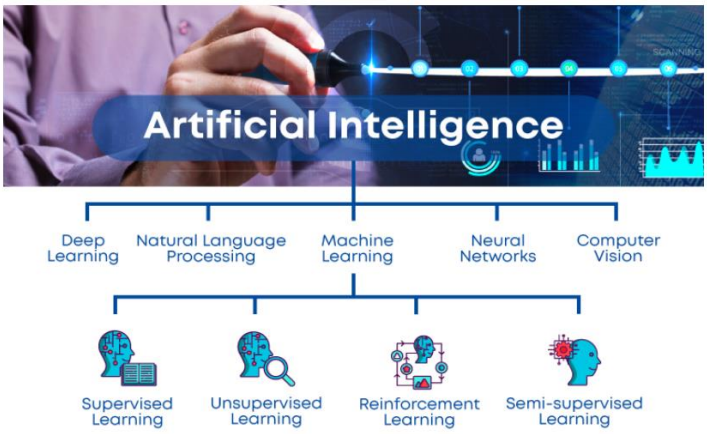
1. الذكاء الإصطناعى:

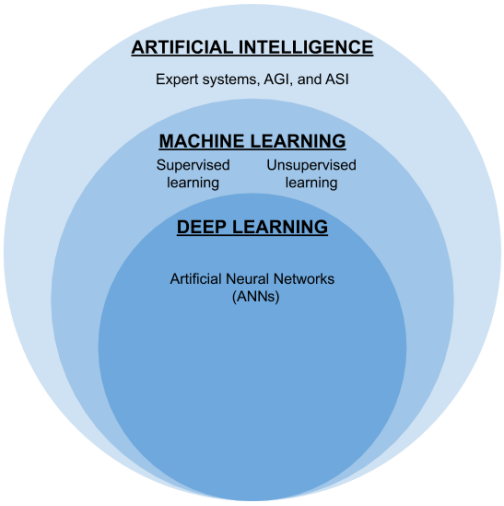
إذا تم تنفيذ المهمة من قبل الإنسان باستخدام الذكاء فإنه يسمى الذكاء الطبيعي. وإذا أضفنا هذا الذكاء البشري الطبيعي إلى آلة يسمى الذكاء الغير طبيعى AI. هنا الآلة ليست مجرد روبوت. ويمكن أن يكون حتى جهاز كمبيوتر أيضا. جون مكارثي، مخترع الذكاء الاصطناعي قدم تعريفه في مؤتمر عام 1956. والتعريف يقول **"علم وهندسة صنع الآلات الذكية"**. على الرغم من أن الذكاء الاصطناعي من علوم الكمبيوترولكن في الوقت الحالي يتم استخدامه من قبل جميع الأقسام مثل المجالات الميكانيكية والسيارات والطيران والمجالات الطبية وما إلى ذلك.



شكل 1: الذكاء الإصطناعى والعلوم الحديثة

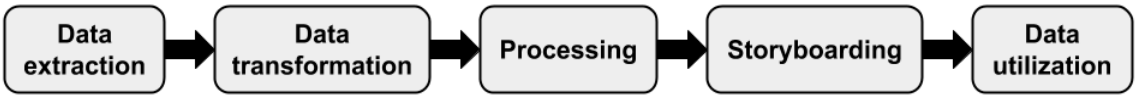
What is AI?





شكل 2 :أ – ب أقسام الذكاء الإصطناعى

و سنلاحظ بعض المعالم الرئيسية التي يمكن للذكاء الاصطناعي من خلالها تمكين برامج التحول الرقمي المستقبلية ومعالجة التحديات الرئيسية. حيث يمثل الرسم البياني جميع المعالم الهامة لمشروع التحول الرقمي فى عصر الذكاء الاصطناعي.

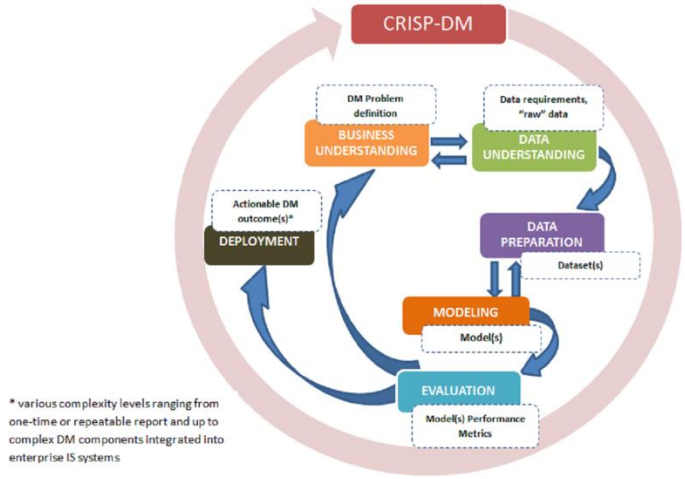


شكل 3: المعالم الهامة في التحول الرقمي باستخدام الذكاء الاصطناعي

ويعتبر التعلم الآلي فرع من فروع الذكاء الاصطناعي الذي يركز على تطوير الخوارزميات والنماذج التي تمكن أجهزة الكمبيوتر من التعلم واتخاذ التنبؤات أو القرارات دون برمجتها بشكل صريح. أنها تنطوي على استخدام التقنيات الإحصائية والخوارزميات لتحليل وتفسير كميات كبيرة من البيانات. حيث يمكن تصنيف خوارزميات التعلم الآلي إلى التعلم الخاضع للإشراف، والتعلم غير الخاضع للإشراف، والتعلم المعزز.

شكل 4: التعلم الآلى

ويتضمن التعلم الخاضع للإشراف تدريب نموذج باستخدام البيانات المصنفة، حيث تكون المخرجات المطلوبة معروفة. ويتعلم النموذج من الأمثلة المسماة لعمل تنبؤات بشأن البيانات الجديدة غير المرئية. ومن ناحية أخرى، يتعامل التعلم غير الخاضع للرقابة مع البيانات غير المسماة. يتعلم النموذج كيفية العثور على الأنماط والهياكل في البيانات دون أي تسميات محددة مسبقًا. وتشمل الأنواع الأخرى من التعلم الآلي التعلم شبه الخاضع للإشراف والتعلم المعزز والتعلم العميق.

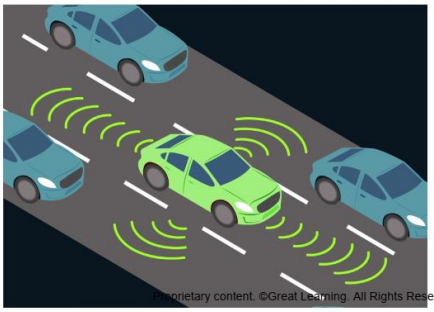


شكل 5: مراحل التعلم الآلى

ويكون التركيز الرئيسي للتعلم الآلي على تطوير الخوارزميات والنماذج لأجهزة الكمبيوتر للتعلم والتنبؤ. إنشاء أنظمة ذكاء اصطناعي تحاكي السلوك البشري.تصميم برامج حاسوبية يمكنها حل المسائل الرياضية المعقدة بناء روبوتات يمكنها أداء المهام دون تدخل بشري. وهناك نطبيقات الذكاء الاصطناعي فى صناعة السيارات كبيرة جدًا حيث تستثمر جميع شركات صناعة السيارات الكبرى مبالغ هائلة من المال لإنشاء سيارة ذاتية القيادة السيارة، والتي يجب أن تعمل بدقة لا تشوبها شائبة تقريبًا لتجنب وقوع الحوادث. في هذا المثال يمكننا أن نرى سيارة Tesla في وضع Auto Pilot.حيث تنتقل السيارة من موقع إلى موقع آخر بشكل تلقائي. يستخدم 3 أجهزة استشعار أمامية تستخدم لتحديد المشاة والمركبات الأخرى بناءً على المسافة بين السيارة والجسم.

Artificial Intelligence is:

- Capable of performing tasks intelligently without being explicitly instructed
- Capable of thinking and acting rationally and humanely



Computer Vision

Video and Traffic analytics

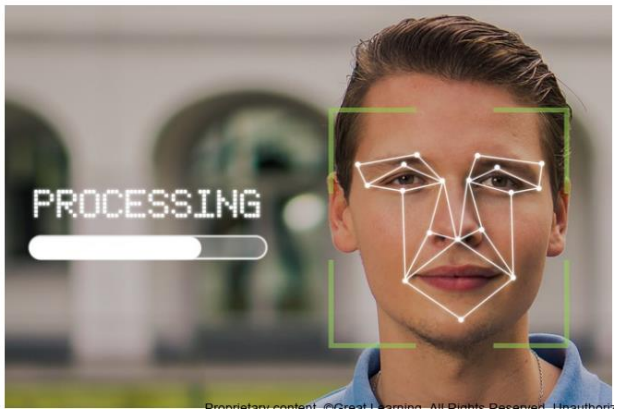
شكل 6: السيارت الذكيه

إنها ضبط السرعة والتحكم في التوجيه وكل شيء. لبناء سيارة ذاتية القيادة ويستخدم مزيجًا من رؤية الكمبيوتر والتعرف على الصور والتعلم العميق. الذي سنفعله أن تتعلمها في المستقبل. تُستخدم نفس التقنيات أيضًا في أنظمة أخرى مثل روبوتات التوصيل والطائرات بدون طيار والمجهولة الهوية. الذكاء الاصطناعي في هذه المجالات له نطاق كبير! الشيء الوحيد هو أنك يجب أن تعرف كيفية التنفيذ في البيئة المعقدة لجعلها خالية من العيوب. المجال التالي الأكثر أهمية هو الذكاء الاصطناعي في قطاع الرعاية الصحية. هذا هو المجال الذي سيكون فيه الذكاء الاصطناعي أكثر فائدة للطاقم الطبي لتقليل ضغط العمل لديهم. لأننا أبدا تعرف على عدد المرضى الذين سيأتون في اليوم الواحد. في الصين لقد قاموا بالفعل بتطبيق الروبوتات المزودة بتقنية الذكاء الاصطناعي في أكثر من 50 مستشفى والتي يتم استخدامها للعملاء غرض الرعاية وللمساعدة في جدولة مواعيد المتابعة أو توجيه المرضى خلال عملية الإستقبال ومساعدي الصحة الافتراضية الذين يقدمون تعليقات طبية أساسية. يقوم الروبوت أيضًا بتسجيل الصوت بين المرضى والأطباء لإنشاء سجل حالة المريض تلقائيًا مما يوفر الكثير من الوقت والجهد للأطباء. مؤخرًا في الصين أجروا مسابقة بين الذكاء الاصطناعي ومجموعة من كبار الأطباء. تم تسليم 225 حالة طبية إلى الذكاء الاصطناعي ومجموعة من 15 طبيبًا من كبار الأطباء لتشخيصها. أورام المخ. في الجولة الأولى، تنبأ الذكاء الاصطناعي بوجود ورم في المخ بنسبة 87% بشكل صحيح خلال 15 دقيقة. في الجولة الثانية. الذكاء الاصطناعي توقعت نسبة 83% بشكل صحيح في 3 دقائق فقط! لكن الأطباء استغرقوا أكثر من 20 دقيقة لفحصها بدقة 66%.هذا لا يعني أن الذكاء الاصطناعي سأحل محل الأطباء! ولكن إذا طبقنا الذكاء الاصطناعي في العيادات، فسيكون الأطباء مجانيين من عبء العمل الثقيل والقدرة على تكريس أنفسهم لتعلم أشياء جديدة وأيضًا سأقوم بتحسين تشخيصهم الكفاءة وتحسين نتائج المريض وتقليل التكلفة.

تطبق الشركات التعلم الآلي لإجراء تشخيص أفضل وأسرع من البشر. واحدة من أفضل تقنيات الرعاية الصحية المعروفة هي IBM Watson. هذه كلها أمثلة قليلة حيث يتم تنفيذ الذكاء الاصطناعي. لدينا العديد من التطبيقات مثل مجرد النظر في الفيسبوك. بمجرد تحميل صورة، يتم تلقائيًا اقتراح علامة الصورة. إنه أيضًا تطبيق للذكاء الاصطناعي باستخدام تقنية معالجة الصور بمساعدة التعرف على الوجه انها تفعل ذلك.

Computer Vision

Face Recognition



شكل 7: التعرف على الوجة

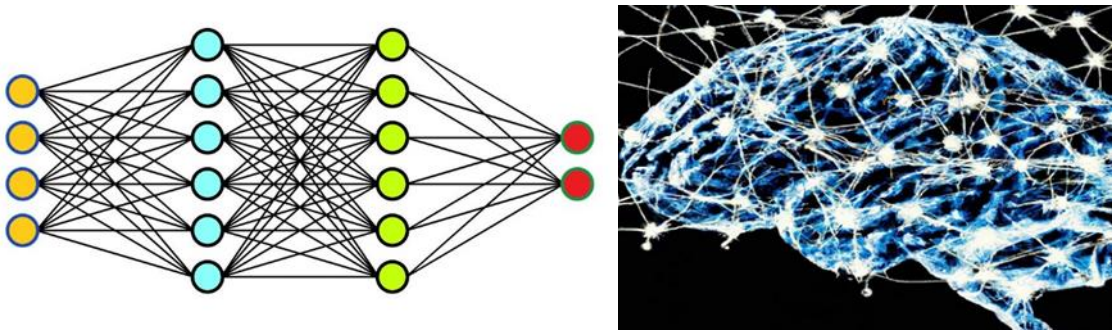
مثال آخر لتوصية المحتوى. توصية الموسيقى على iTunes وتوصية الأفلام على Netflix وتوصية الكتب الإلكترونية في Amazon Kindle.بمساعدة الخوارزمية، يقترح المحتوى بناءً على بياناتك المستخدمة سابقًا

والذكاء الاصطناعي هو على النقيض من الذكاء البشري,و يشير الذكاء الاصطناعي إلى أن الآلات يمكنها تقليد البشر في: تتحدث • التفكير• تعلُّم• تخطيط• فهم يُطلق على الذكاء الاصطناعي أيضًا اسم الذكاء الآلي وذكاء الكمبيوتر. الذكاء الاصطناعي هو نظام علمي يشمل العديد من مجالات علوم البيانات التي تتراوح من الذكاء الاصطناعي الضيق إلى الذكاء الاصطناعي القوي، بما في ذلك التعلم الآلي والتعلم العميق والبيانات الضخمة واستخراج البيانات. التعلم الآلي هو مجال فرعي من الذكاء الاصطناعي "تعلم الآلات لتقليد الذكاء البشري" و الذكاء الاصطناعي الهدف الضيق التعلم الآلي الشبكات العصبية البيانات الكبيرة التعلم العميق الذكاء الاصطناعي القوي

الشبكات العصبية (NN):

الشبكات العصبية هي: • تقنية البرمجة • إحدى الطرق المستخدمة في التعلم الآلي • برنامج يتعلم من الأخطاء وتعتمد الشبكات العصبية على كيفية عمل الدماغ البشري:حيث تقوم الخلايا العصبية بإرسال رسائل لبعضها البعض. فبينما تحاول الخلايا العصبية حل مشكلة ما (مرارًا وتكرارًا)، فإنها تعمل على تقوية الروابط التي تؤدي إلى النجاح وتقليل الروابط التي تؤدي إلى الفشل. الشبكات العصبية هي إدراكات متعددة الطبقات. تتكون الشبكة العصبية في أبسط صورها من:

• طبقة الإدخال (الصفراء) • طبقة مخفية (زرقاء) • طبقة الإخراج (حمراء). في نموذج الشبكة العصبية، تتم معالجة البيانات المدخلة (الصفراء) مقابل طبقة مخفية (الزرقاء) قبل إنتاج المخرجات النهائية (الحمراء). الطبقة الأولى: تتخذ الإدراك الحسي الأصفر قرارات بسيطة بناءً على المدخلات. يتم إرسال كل قرار إلى الإدراك الحسي في الطبقة التالية.و الطبقة الثانية: تتخذ أجهزة الإدراك الحسي الزرقاء قراراتها من خلال وزن النتائج من الطبقة الأولى. تتخذ هذه الطبقة قرارات أكثر تعقيدًا على مستوى أكثر تجريدًا من الطبقة الأولى. لشبكات العصبية العميقة الشبكات العصبية العميقة هي: • تقنية البرمجة • إحدى الطرق المستخدمة في التعلم الآلي • برنامج يتعلم من الأخطاء تتكون الشبكات العصبية العميقة من عدة طبقات مخفية من الشبكات العصبية التي تؤدي عمليات معقدة على كميات هائلة من البيانات. تستخدم كل طبقة متتالية الطبقة السابقة كمدخل. على سبيل المثال، تستخدم القراءة البصرية طبقات منخفضة لتحديد الحواف، وطبقات أعلى لتحديد الحروف.في نموذج الشبكة العصبية العميقة، تتم معالجة بيانات الإدخال (الصفراء) مقابل طبقة مخفية (الأزرق) وتعديلها مقابل المزيد من الطبقات المخفية (الأخضر) لإنتاج المخرجات النهائية (الأحمر). الطبقة الأولى: تتخذ الإدراك الحسي الأصفر قرارات بسيطة بناءً على المدخلات. يتم إرسال كل قرار إلى الإدراك الحسي في الطبقة التالية. الطبقة الثانية: تتخذ أجهزة الإدراك الحسي الزرقاء قراراتها من خلال وزن النتائج من الطبقة الأولى. تتخذ هذه الطبقة قرارات أكثر تعقيدًا على مستوى أكثر تجريدًا من الطبقة الأولى. الطبقة الثالثة: يتم اتخاذ قرارات أكثر تعقيدًا بواسطة الإدراك الحسي الأخضر.



شكل 8: الخلايا العصبية

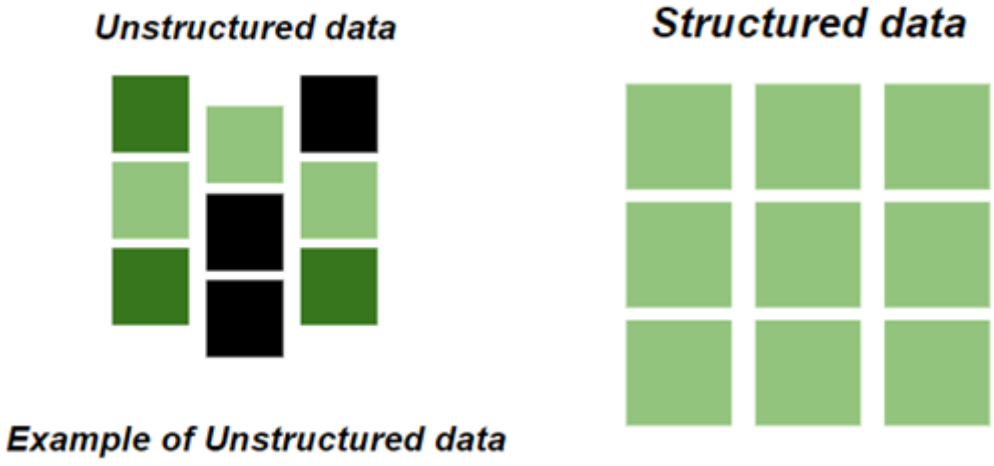
_التعلم العميق (DL):

التعلم العميق هو مجموعة فرعية من التعلم الآلي.التعلم العميق هو المسؤول عن ازدهار الذكاء الاصطناعي في السنوات الأخيرة. التعلم العميق هو نوع متقدم من تعلم الآلة الذي يتعامل مع المهام المعقدة مثل التعرف على الصورة .

2. علم البيانات

هو مزيج من التخصصات المتعددة التي تستخدم الإحصائيات وتحليل البيانات والتعلم الآلي لتحليل البيانات واستخلاص المعرفة والرؤى منهاوعلم البيانات يدور حول جمع البيانات وتحليلها واتخاذ القرار. يدور علم البيانات حول العثور على أنماط في البيانات، من خلال التحليل، وعمل تنبؤات مستقبلية. باستخدام علم البيانات، تستطيع الشركات القيام بما يلي: قرارات أفضل (هل يجب أن نختار A أو B) و حيث التحليل التنبؤي (ماذا سيحدث بعد ذلك؟)اكتشافات الأنماط (ابحث عن النمط، أو ربما المعلومات المخفية في البيانات) يُستخدم علم البيانات في العديد من الصناعات في العالم اليوم، على سبيل المثال. الخدمات المصرفية والاستشارات والرعاية الصحية والتصني.

و أحد أهداف علم البيانات هو تنظيم البيانات، مما يجعلها قابلة للتفسير وسهلة التعامل معها. يمكن تصنيف البيانات إلى مجموعتين •: البيانات المنظمة • البيانات غير المنظمة البيانات غير المنظمة البيانات غير المنظمة ليست منظمة. يجب علينا تنظيم البيانات لأغراض التحليل .كيفية هيكلة البيانات: يمكننا استخدام مصفوفة أو جدول قاعدة بيانات لتنظيم البيانات أو تقديمها. مثال على المصفوفة: [80، 85، 90، 95، 100، 105، 110، 115، 120، 125] يوضح المثال التالي كيفية إنشاء مصفوفة في بايثون

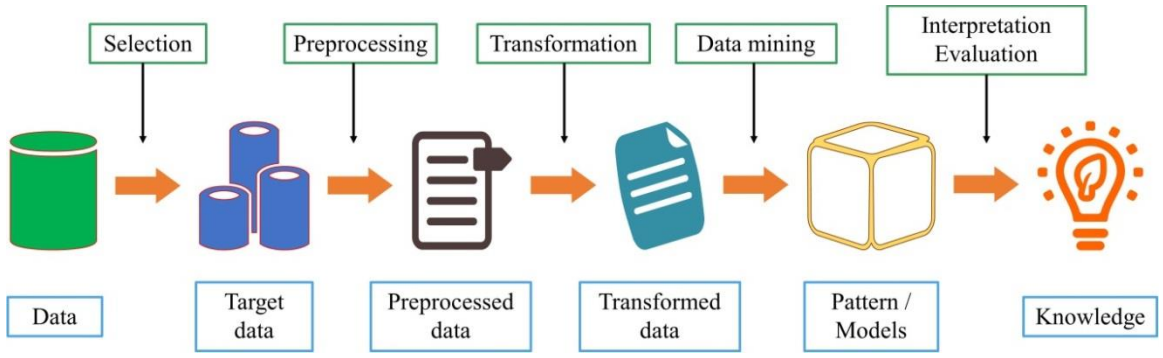


شكل 9: أنواع البيانات

وبعرف علم البيانات بأنه مجال متعدد التخصصات يجمع بين التقنيات والأدوات والمنهجيات المختلفة لاستخراج رؤى ومعرفة قيمة من البيانات. إنها تنطوي على عملية جمع وتنظيف وتحليل وتفسير كميات كبيرة من البيانات

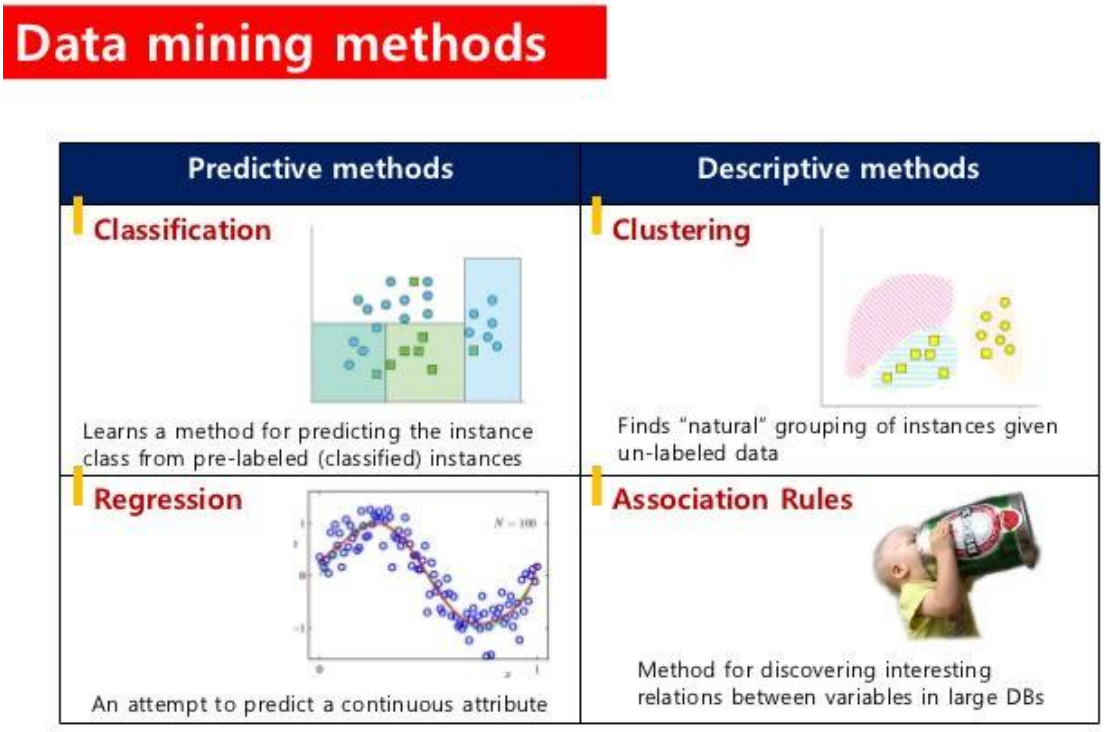
المنظمة وغير المنظمة لاتخاذ قرارات مستنيرة وحل المشكلات المعقدة. ومع النمو الهائل للبيانات في السنوات الأخيرة، أدركت المؤسسات عبر الصناعات أهمية علم البيانات في دفع الابتكار وتحسين العمليات واكتساب ميزة تنافسية. يلعب علماء البيانات دورًا حاسمًا في إطلاق إمكانات البيانات من خلال تطبيق التقنيات الإحصائية والحسابية، وخوارزميات التعلم الآلي، وأدوات تصور البيانات. و استكشاف مجموعات البيانات وتحديد الأنماط وتطوير النماذج التنبؤية لاستخراج معلومات ذات معنى. و ستطبق المهارات التحليلية للكشف عن الرؤى الخفية التي يمكن أن تقود استراتيجيات العمل، وتحسين العمليات، وتعزيز عملية صنع القرار.

يشمل علم البيانات مجموعة واسعة من التقنيات والأساليب، بما في ذلك التحليل الإحصائي، والتعلم الآلي، واستخراج البيانات، والنمذجة التنبؤية، وتصور البيانات، والمزيد. ويتطلب إتقان لغات البرمجة مثل Python أو R، بالإضافة إلى المعرفة بقواعد البيانات ومعالجة البيانات وتقنيات تنظيف البيانات.

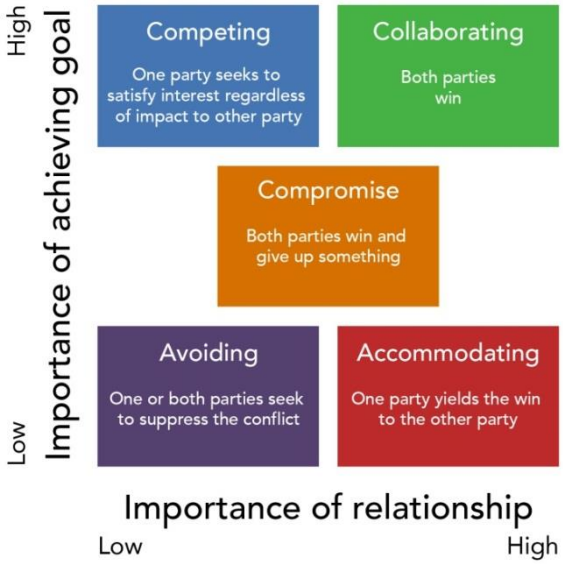


شكل 10: مراحل علم البيانات

وتعتبر المفاهيم والتقنيات الأساسية لعلم البيانات، والمهارات اللازمة للتفوق في هذا المجال. مثل استكشاف البيانات، والمعالجة المسبقة للبيانات، والتحليل الإحصائي، وخوارزميات التعلم الآلي، وتصور البيانات. ويعد علم البيانات مجالًا متناميًا أحدث ثورة في الصناعات في جميع أنحاء العالم. إنها ممارسة استخلاص الأفكار والمعرفة من كميات هائلة من البيانات باستخدام الأساليب والخوارزميات والأدوات العلمية المختلفة. في عالم اليوم القائم على البيانات، يلعب علم البيانات دورًا حاسمًا في تشكيل استراتيجيات الأعمال، وتحسين عمليات صنع القرار وتحفيز الإبتكار



شكل 11: استخلاص المعلومات



شكل 12: العلاقة بين الأهميه والأولويات

أحد الأدوار الأساسية لعلم البيانات في الصناعة هو التحليلات التنبؤية. تتضمن التحليلات التنبؤية استخدام البيانات التاريخية وتقنيات النمذجة الإحصائية للتنبؤ بالنتائج المستقبلية. من خلال تحليل الأنماط والاتجاهات في البيانات، يمكن لعلماء البيانات التنبؤ بسلوك العملاء واتجاهات السوق والمخاطر المحتملة. يتيح ذلك للشركات اتخاذ قرارات مستنيرة وتطوير استراتيجيات للبقاء في صدارة المنافسة.

يلعب علم البيانات أيضًا دورًا مهمًا في تحسين العمليات التجارية. من خلال استخدام تقنيات مثل استخراج البيانات والتعلم الآلي، يمكن لعلماء البيانات تحديد أوجه القصور والاختناقات في العمليات واقتراح التحسينات. من خلال تحليل البيانات من مصادر مختلفة، يمكن لعلماء البيانات اكتشاف الأنماط والعلاقات المخفية التي يمكن أن تؤدي إلى تخصيص أكثر كفاءة للموارد، وخفض التكاليف، وزيادة الإنتاجية.

علاوة على ذلك، فإن علم البيانات له دور فعال في تعزيز تجربة العملاء. من خلال تحليل بيانات العملاء، مثل سجل الشراء وسلوك التصفح والتعليقات، يمكن للشركات الحصول على رؤى قيمة حول تفضيلات العملاء واحتياجاتهم. يتيح لهم ذلك تخصيص المنتجات والخدمات وتصميم الحملات التسويقية وتقديم خدمة عملاء استثنائية. ومن خلال الاستفادة من علم البيانات، يمكن للشركات إنشاء نهج شخصي ومستهدف يعزز رضا العملاء وولائهم.

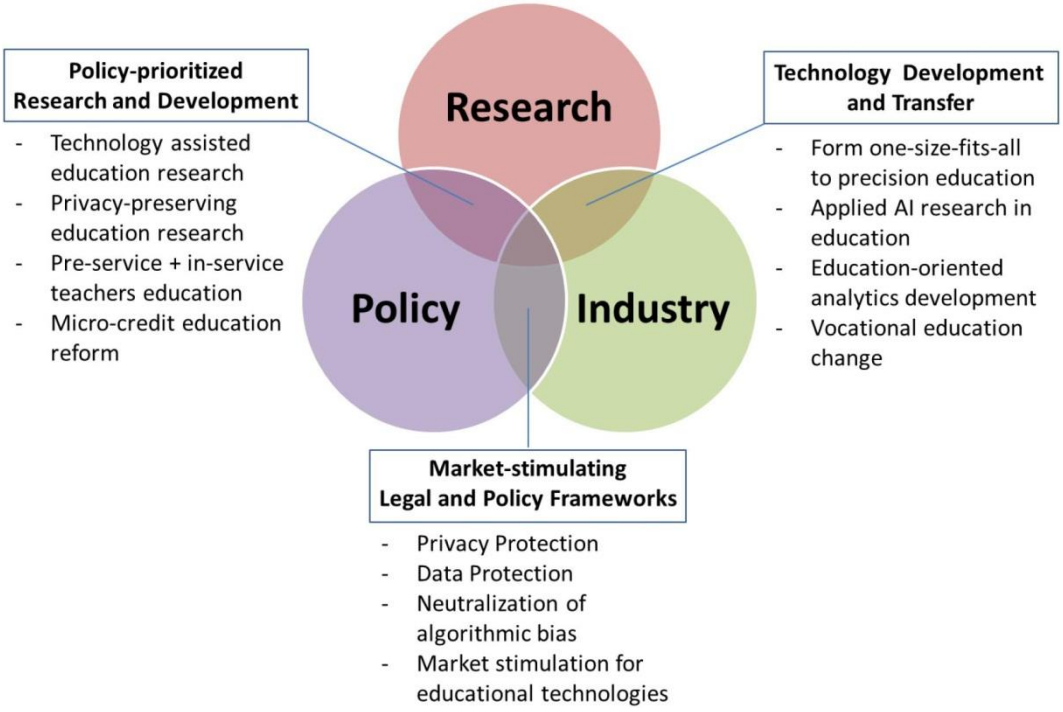
هناك دور حيوي آخر لعلم البيانات في الصناعة وهو مجال إدارة المخاطر. ومن خلال تحليل البيانات التاريخية واستخدام تقنيات النمذجة المتقدمة، يمكن لعلماء البيانات تقييم المخاطر وتخفيفها. وهذا مهم بشكل خاص في صناعات مثل التمويل والتأمين، حيث يمكن لتقييم المخاطر الدقيق أن يمنع الخسائر المالية ويحسن عمليات صنع القرار.

فلقد ، أصبح علم البيانات جزءًا لا يتجزأ من الصناعة. من التحليلات التنبؤية إلى تحسين العمليات التجارية وتعزيز تجربة العملاء، يتمتع علم البيانات بالقدرة على تحويل الأعمال ودفع الابتكار. مع استمرار نمو كمية البيانات بشكل كبير، سيصبح دور علم البيانات أكثر أهمية في تشكيل مستقبل الصناعات في جميع أنحاء العالم.

3. البيانات الضخمة، والثورة الصناعية الرابعة ، وعلم البيانات:

كانت البيانات الضخمة والتقدم السريع في تقنيات الحوسبةالى تحول مجتمعنا بسرعة بالتحول الرقمي ويدخل إلى عالم جديد مع العديد من التطورات المذهلة. ويوجد العديد من المفاهيم، والتقنيات، والأدوات، والتطبيقات الجديدة التي يحتاج طلابنا إلى تعلمها مثل: Hadoop / MapReduce، وSpark، وNoSQL،

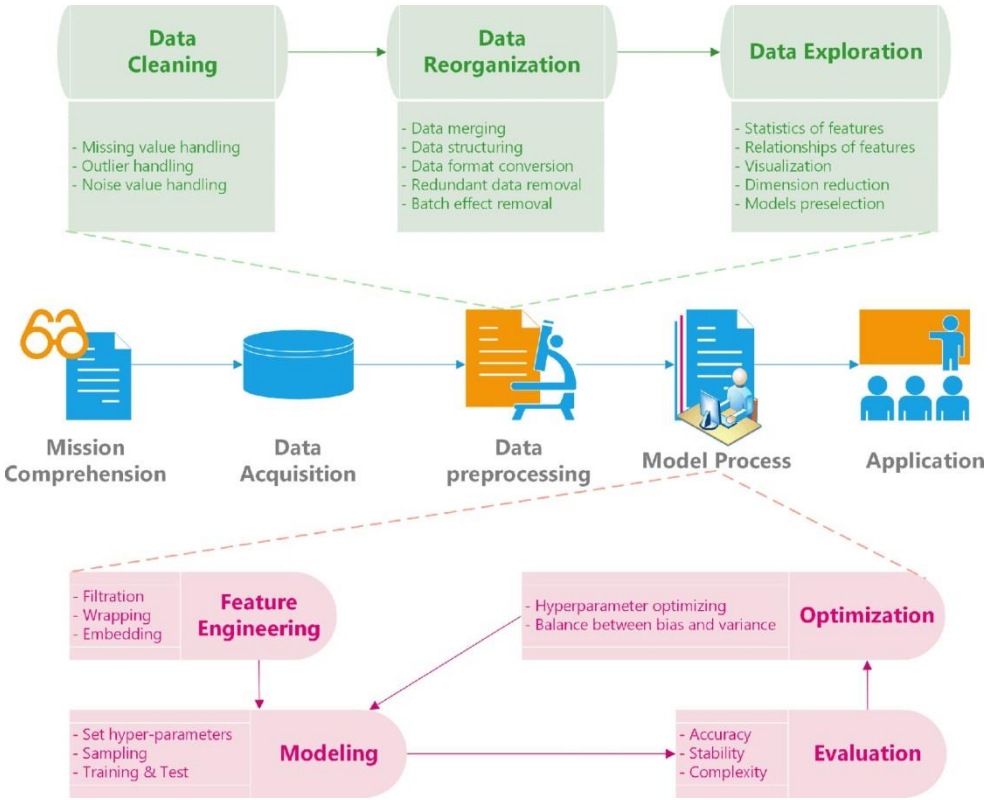
و NewSQL، والحوسبة في الذاكرة، وتمثيل البيانات، واختزان البيانات الضخمة، وبحيرة البيانات data lake، والحوسبة السحابية، وإنترنت الأشياء (IoT)، والذكاء الاصطناعي، والروبوتات التي تتواصل مع الواقع البشري والواقع الافتراضي، والواقع المعزز، والتعلم الآلي، والتعلم العميق، والحوسبة الإدراكية، وتحليلات البيانات الضخمة. لقد ساهمت هذه التقنيات في ازدهار الثورة الرقمية من خلال تطوير العديد من التطبيقات المبتكرة في جميع القطاعات الصناعية والمجتمعية. وعلى الرغم من أن طلاب كليات المعلومات ليسوا مضطرين لتعلم الجوانب التقنية لجميع هذه التقنيات، إلا أنهم بحاجة إلى استيعاب المفاهيم، وكذلك أدوارهم ونقاط قوتهم وحدودهم. كما أنهم بحاجة إلى تعلم كيفية إنشاء تطبيقات جديدة استنادًا إلى الأساليب الحالية، وتعلم الأدوات التي تدعم تلك التقنيات، مما يتيح لهم التفاعل مع التطبيقات.



شضكل 13: **البيانات الضخمة، والثورة الصناعية الرابعة ، وعلم البيانات**

وتكمن الحركة الملاحظة في هذه التطبيقات المبتكرة في انصهار التقنيات المتعددة في التطبيقات الفيزيائية، والميكانيكية، والبيولوجية. على سبيل المثال، فكر في سيناريو يوظف وبشكلٍ فعال تقنيات الشيخوخة الذكية التي تخدم السكان المسنين الذين تتزايد أعدادهم بسرعة. ويتم الآن جمع أنواع مختلفة من البيانات الصحية والبيانات الموقفية الخاصة بالمسنين بواسطة أجهزة الاستشعار القابلة للارتداء، وأجهزة استشعار إنترنت الأشياء المثبتة في المنازل والمستشفيات. ويمكن تخزين البيانات الخام التي يتم جمعها في السحابة على شكل بحيرة بيانات. وتفرض الميتاداتا الدلالية ضرورة استخراجها من البيانات الأولية بطريقةٍ شبه تلقائية، متى تم تحديد الأحداث

والسيناريوهات المثيرة للاهتمام، كما أن مخطط بيانات سجل الحياة يحتاج إلى البناء. وخلال هذه العملية ، يتم استخدام خوارزميات التعلم الآلي للتعرف على سلوك المسنين. يتم بناء ملف شخصي لكبار السن من خلال دمج البيانات المستفادة وبيانات الرعاية الصحية الإلكترونية. قد يتواصل الروبوت الذكي للرعاية البشرية مع كبار السن من خلال التعلم من ملفاتهم الشخصية والبيانات المتكاملة من مصادر مختلفة. يمكن تحديد أي أعراض أو سلوكيات غير طبيعية لكبار السن من البيانات المدمجة التي تم دمجها من البيانات المستفادة ، وبيانات المشاعر المستخرجة من وسائل التواصل الاجتماعي ، والبيانات في الوقت الحقيقي التي يتم تغذيتها من أجهزة الاستشعار القابلة للارتداء وأجهزة استشعار إنترنت الأشياء. يستخدم أخصائيو الرعاية الصحية لوحة أجهزة القياس التي تم إنشاؤها باستخدام أدوات تحليل البيانات الضخمة والتواصل مع كبار السن لأية رسائل مهمة عبر الهاتف الذكي. كما هو موضح في هذا المثال ، يستخدم التطبيق المبتكر مزيجًا من التقنيات المتعددة. من الممكن إنشاء خدمة أو نشاط تجاري جديد من خلال جمع البيانات من خلال مستشعرات إنترنت الأشياء ، وتخزينها في السحابة ، وتحليلها من خلال الذكاء الاصطناعي وتحليلات البيانات الضخمة.

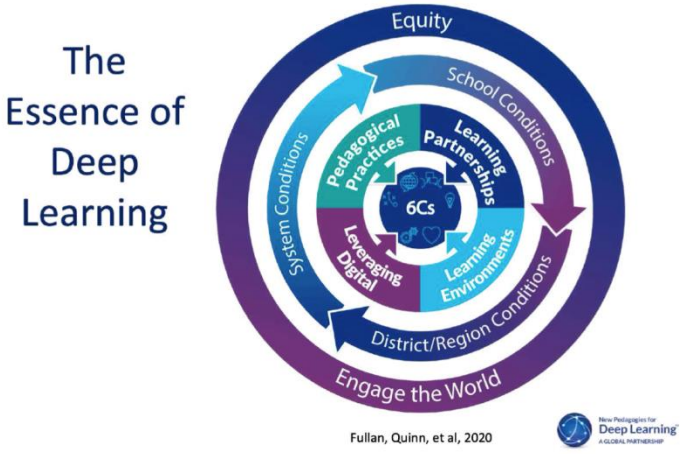


شكل 14: تحليل البيانات وتوظيفها

ويُطلق على النموذج الحديث، الذي يشرح الصناعة فائقة الترابط super-connected industry، التي **تدمج تقنيات متعددة للأغراض التصنيعية** مصطلح الثورة الصناعية الرابعة. وتؤكد الصناعة 4.0 على عالم

ذكي ومتصل بالشبكة من خلال تقديم مفهوم النظم السيبرانية الفيزيائية Cyber–Physical Systems (CPS) ، والتي تتضمن "الأجهزة الذكية ونظم التخزين ومرافق الإنتاج القادرة على تبادل المعلومات بشكلٍ مستقلٍ وتحريك الإجراءات والتحكم في بعضها البعض بشكلٍ مستقل"

وتعتبر التأثيرات الواسعة للثورة الصناعية الرابعة على الأعمال والحكومة والناس. وتتلخص هذه التأثيرات بإيجاز فيما يأتي: (1) سيتم تزويد العملاء بالمنتجات والخدمات ذات القيم المتزايدة بسبب التقنيات والبيانات والتحليلات الجديدة (جانب الأعمال)، (2) سيتم تمكين المواطنين بشكلٍ متزايد من التعامل مع الحكومات في حين سيكون لدى الحكومات المزيد من الأدوات لزيادة متابعتها على السكان (3) ستتأثر مفاهيم الخصوصية والملكية ونمط الاستهلاك والوظائف وتنمية المهارات. وبشكل عام، ترتبط جميع التأثيرات المذكورة أعلاه ارتباطًا وثيقًا بربط الأشخاص بالقيم. وتتميز العواقب الاقتصادية العالمية المحتملة بالأتمتة، والتواصلية الفائقة ومن المتوقع أن يستمر أتمتة الوظائف ذات المهارات المنخفضة والمتوسطة إلى أن تختفي في نهاية الأمر. وتعد السيارات بدون سائق والطائرات بدون طيار خير مثال لهذا الاتجاه. وسوف تكون إنترنت الأشياء وقضايا الأمن السيبراني Cybersecurity ذات الصلة من أبرز ما تنتجه الثورة الصناعية الرابعة ومن المحتمل بشكلٍ كبيرٍ أن تكمن المشكلات الأكثر تعقيدًا في الجانب السيبراني للنظم السيبرانية الفيزيائية بدلاً من الجانب المادي التقليدي، الأمر الذي يتطلب منا أن نكون مجهزين بعقلية جديدة حول كيفية رؤيتنا للعالم وإدارة للمواردنا.



شكل 15: النموذج الحديث، الذي يشرح الصناعة فائقة الترابط

وفي الثورة الصناعية الرابعة ، تضطلع البيانات الضخمة وتقنياته ولا سيما الذكاء الاصطناعي الذي يتضمن التعلم الآلي والروبوتات، أدوارًا مهمة. ويعتقد الخبراء أنه سيكون لها تأثيرات كبيرة على حياتنا المستقبلية بما

يساعدنا على التعامل مع التحديات الجديدة. وباتت هذه التحديات والفرص الجديدة وشيكة بشكلٍ متزايدٍ، وعلينا أن نفهمها وأن نستعد لها من أجل الاستفادة الكاملة من مزاياها. ويوجد العديد من الطرق التي يمكننا من خلالها الاستعداد للثورة الصناعية الرابعة ، مثل دمج الشركات غير المتصلة بالإنترنت مع الشركات عبر الإنترنت، ودمج الواقع الافتراضي والعالم الحقيقي لتطوير الخدمات والصناعات التي تركز على الإنسان.

وتؤثر هذه الاتجاهات الحديثة افتراضياً على جميع المجتمعات والوظائف المستقبلية، بما في ذلك وظائف الطلاب، الذين يجب أن يستعدوا للتحديات الرقمية الجديدة وعلم البيانات التي تتضمن بيانات ضخمة. ومن أجل تخريج المواهب المناسبة لهذه البيئات الرقمية الجديدة والاحتياجات التعليمية المستقبلية، سيكون من الأفضل تعليم طلاب كليات المعلومات في إطار برامج علم البيانات المزدهرة حديثًا. ويمكن أن يكون علم البيانات هو التخصص الرئيسي الذي يدعم الجوانب الرئيسية لمختلف هذه التطورات الجديدة. لقد أصبح علم البيانات أحد التخصصات الأكثر جاذبية في السنوات الأخيرة، حيث أنه يوفر العديد من الفرص للطلاب والعلماء الموهوبين. وتبحث القطاعات الصناعية عن علماء بيانات الماهرين والموهوبين في تقنيات البيانات الضخمة والقادرين على تطبيق تخصصات علم البيانات في المجالات المختلفة باستخدام أدوات حل المشكلات المناسبة. كما بدأت العديد من الجامعات في تقديم برامج تعليمية لعلم البيانات على جميع مستويات الدراسة الجامعية، والماجستير، ودكتوراه. وفي قلب مجال المعلومات هذا، نجحت كليات المعلومات في تخريج علماء المعلومات الذين يجيدون إجراء الدراسات/ العمل المتعلق بتجميع البيانات، والمعلومات، والمعرفة. وقد حان الوقت الآن لتقوم مدارس المعلومات بتعزيز علماء البيانات الموهوبين من خلال الاستفادة من نقاط القوة في تخصصات علم المعلومات.

ومع وجود العديد من التحديات الرقمية الناشئة سريعة النمو، تناقش هذه المقالة وجهات النظر حول فرص كليات المعلومات والاقتراحات الخاصة بتعليم علم البيانات. ويتعين على كليات المعلومات أن تمكن طلابها من خلال تخصصات "حوسبة المعلومات"، التي نعرّفها على أنها القدرة على حل المشكلات وإنشاء القيم والمعلومات والمعرفة باستخدام الأدوات في المجالات التطبيقية المختلفة. وباعتبارها مناهج متخصصة لدعم تخصصات حوسبة المعلومات في تعليم علم البيانات، فنقترح المحاور الثلاثة القائمة على **المستخدم، والأداة، والتطبيق.** ستعمل هذه المحاور الثلاثة على التمييز بين تعليم علم البيانات في مدارس المعلومات، وتعليم علوم الحاسب أو كليات إدارة الأعمال. ونقدم إطار تعليم علم البيانات متعدد الطبقات Data Science Education Framework (DSEF)، ذو وحدات بناء، تتضمن الركائز الثلاث لعلم البيانات (الأشخاص، والتقنية، والبيانات)، والتفكير الحاسوبي، والنماذج القائمة على البيانات، ودورات حياة علم البيانات. وبالتالي ، يجب تصميم مقررات علم البيانات استناداً إلى هذا الإطار باستخدام المناهج القائمة على المستخدم، والقائمة على

الأدوات والقائمة على التطبيق. وسوف يساعد هذا الإطار طلابنا على التفكير في مشكلات علم البيانات من منظور الصورة الكبيرة وتعزيز مهارات حل المشكلات المناسبة في مختلف مراحل دورة حياة علم البيانات.

4. تعليم علم البيانات وكليات المعلومات:

4-1- عالم البيانات:

إن مصطلح "عالم البيانات" قد صيغ في عام 2008م ومن التعريفات الشائعة لعلم البيانات أنه "**دراسة الاستخراج المعرفة القابل للتعميم من البيانات**" ، وأنه "مجال عمل ناشئ معني **بجمع، وإعداد، وتحليل، وتمثيل، وإدارة، وحفظ مجموعات كبيرةٍ من المعلومات**". إن "علم البيانات يتضمن المبادئ، والعمليات، والتقنيات التي تساعد في فهم الظواهر من خلال التحليل (الآلي) للبيانات." إن "علماء البيانات هم مهنيون رفيعو المستوى، لديهم التدريب والفضول الذي يمكنهم من باستكشاف عالم البيانات الضخمة، فهم يضطلعون بعمليات الاستكشاف أثناء السباحة في البيانات، ويقومون بإيصال ما تعلموه، ويقدمون المقترحات الخاصة بتأثير تداعيات النتائج على اتجاهات الأعمال الجديدة ". وبناءً على هذه التعريفات، يوجد ثلاثة أبعاد مهمة لعلماء البيانات. يجب أن يكون علماء البيانات قادرين على:

(1) فهم أدوار البيانات الضخمة وتقنيات البيانات الضخمة،

(2) العمل مع جميع مراحل دورة حياة علم البيانات – اكتشاف المشكلات، وحل المشكلات، وإيصال الحلول،

(3) استخدام مجموع من الأدوات لحل مشاكل البيانات الضخمة. ونناقش في الأقسام التالية كيف يجب أن تتعامل كليات المعلومات مع تعليم علم البيانات.

4-2- التحديات في تعليم علم البيانات:

تكافح الشركات الصناعية لتوظيف علماء البيانات الموهوبين، للتعامل مع متطلبات معالجة البيانات المتزايدة التعقيد والتحديات ذات الصلة. وعلى الرغم من أن العديد من الجامعات قد بدأت في تقديم برامج تعليم علم البيانات على مستويات مختلفة، إلا أن تقنيات البيانات الضخمة تعد من الموضوعات الجديدة نسبيًا في مناهج التخصصات التقليدية. لقد اكتسب معظم علماء البيانات مهارات التعامل مع البيانات الضخمة خارج أسوار الجامعة. ونظرًا لنطاقها الواسع، يشكل دمج تقنيات البيانات الضخمة بسلاسة في المناهج تحديًا كبيرًا للأقسام التعليمية التقليدية. وبشكل عام، يمكن تلخيص تلك التحديات على النحو التالي:

(1) كيفية دمج ثورة البيانات الضخمة في تعليم علم البيانات بشكلٍ سلس،

(2) كيفية تعليم الطلاب تصميم الأدوات، أو القيم الخاصة بالمعلومات، أو المعارف معرفة من البيانات الضخمة من خلال توظيف من مهارات علم البيانات،

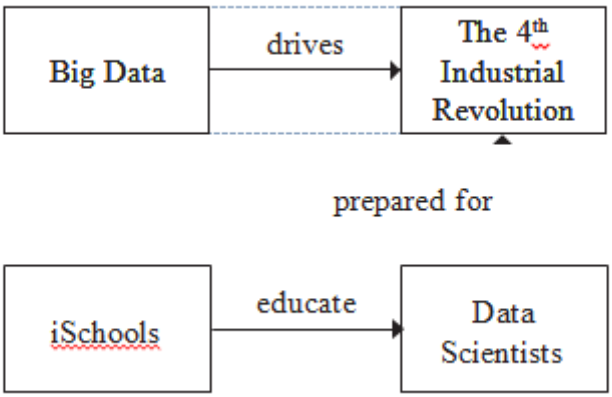
(3) كيفية تطوير المناهج الدراسية لتستوعب أحدث الموضوعات مثل: التعزيز الذكي intelligent augmentation، والحوسبة الادراكية cognitive computing،

(4) كيفية تسكين موضوعات إدارة البيانات الناشئة هذه في المناهج الدراسية.

4-3- كليات المعلومات المستقبليه كمراكز لتعليم علم البيانات:

تعدكليات المعلومات واحدة من المجتمعات الرائدة في تعليم علم البيانات. لتطوير مجال المعلومات والذهاب به فيما وراء علم المكتبات، وربط طرفي البيانات والمعرفة. وعمدت كليات المعلومات العصرية إلى تدريب طلابها كباحثين، ومتخصصين متنوعين يمتلكون مهارات متعددة الجوانب لإدارة المعلومات وتحليل البيانات. وقد يرجع السبب في ذلك إلى تعدد التخصصات التي يجب أن تتمتع بها كليات المعلومات، والتي تتأتى من خلال تعيين أعضاء هيئة تدريسية ينتمون إلى خلفيات تعليمية متنوعة، مثل تضمين علوم المكتبات والمعلومات، وعلوم الحاسب، والتعليم، والاتصالات، والإدارة، والتاريخ، ونظم المعلومات، واللغة الإنجليزية، وعلم النفس، والفلسفة، والهندسة الصناعية . ولسوف تمكّن أصول الموارد البشرية متعددة التخصصات كليات المعلومات من إعداد علماء بيانات ناجحين ومسلحين بمهارات متنوعة، على عكس الأقسام الأخرى التي تمتلك هيئة تدريسية ترتكز خبراتها في مجال واحد. وعلى ذلك، ربما تكون **كليات المعلومات المستقبليه** هى المؤسسات الأكثر مثالية لتدريس لتعليم علم البيانات الذي يعتمد في تدريسه على المستخدم **teaching user-based**، ويركز على التطبيق.

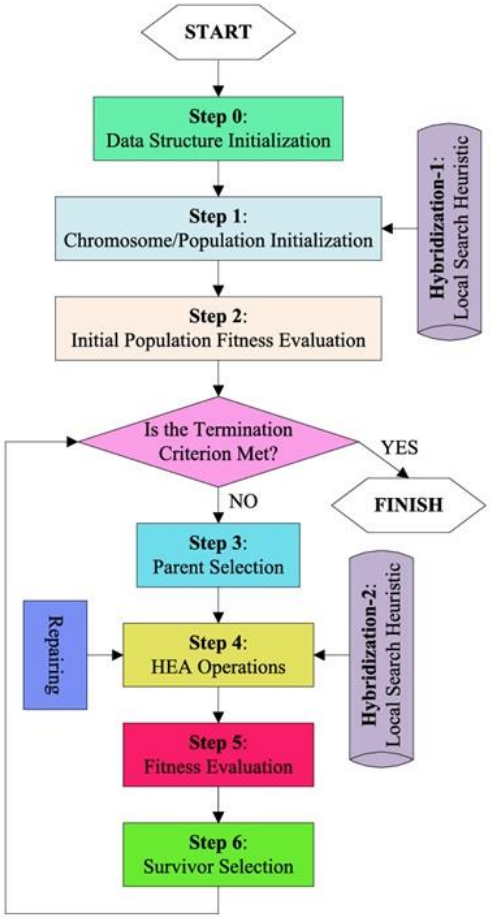
وباختصار، تقود البيانات الضخمة الثورة الصناعية الرابعة، ويتعين على **كليات المعلومات المستقبليه** إعداد علماء البيانات الناجحين القادرين على مواجهة هذه الثورة.



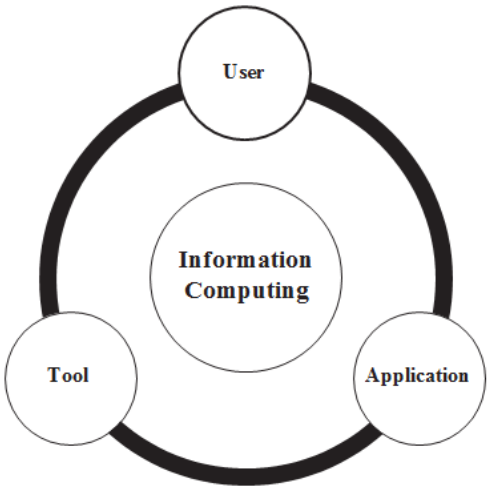
شكل 16: كليات المعلومات المستقبليه تقوم بتعليم علماء البيانات لإعدادهم للثورة الصناعية الرابعة.

5. محاور تعليم علم البيانات في كليات المعلومات المستقبليه:

يمتلك كل قسم رؤيته وطرق تدريس علم البيانات الخاصة به. بالنسبة إلى **كليات المعلومات المستقبليه** ، فقد يكون مفهوم حوسبة المعلومات، الذي يُعرف بأنه القدرة على حل المشكلات وإنشاء القيم، والمعلومات، والمعرفة باستخدام الأدوات في مجالات التطبيق هو الأكثر ملاءمةً. وتختلف حوسبة المعلومات في **كليات المعلومات المستقبليه** عن الحوسبة التقليدية في علوم الحاسب. تهتم علوم الحاسب بتدريس الحوسبة بمزيد من التركيز على العناصر الأساسية مثل: هياكل البيانات، والخوارزميات، والنظرية الحاسوبية، ونماذج الحوسبة. ومن ناحية أخرى، يجب أن تركز حوسبة المعلومات في مدارس المعلومات على المستخدمين، والأدوات، والتطبيقات. ومن أجل تنفيذ حوسبة المعلومات لتعليم علم البيانات في مدارس المعلومات، نوصي بالمحاور الثلاثة: المستندة إلى المستخدم، والمستندة إلى الأدوات، والمستندة إلى التطبيق (الشكل 17)، حيث تحتفظ كلٌ من هذه المحاور برصيد متساوٍ في إنتاج علماء بيانات متوازنين ومتكاملين.



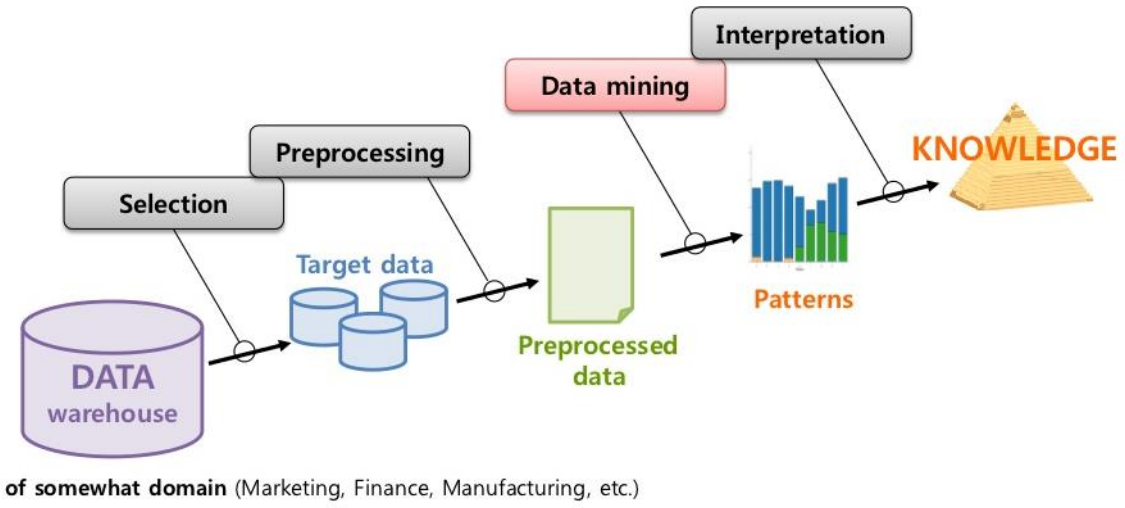
شكل 17: المحاور الثلاثة: المستندة إلى المستخدم، والمستندة إلى الأدوات، والمستندة إلى التطبيق



شكل 18: المحاور الثلاث لحوسبة المعلومات بكليات المعلومات المستقبليه.

5-1 تعليم علم البيانات المستند إلى المستخدم:

يعني تعليم علم البيانات المستند إلى المستخدم تدريب عالم البيانات الذي يحل مشكلات علم البيانات في بعض مجالات التطبيق، أو يمكنه استخدام منتجات علم البيانات بشكلٍ فعال، بدلاً من أولئك الذين يتوفرون على تطوير منتجات علم البيانات. ويتعين على **كليات المعلومات المستقبليه** التركيز على تعليم حل مشكلات علم البيانات على المستويات التطبيقية، بدلاً من المطورين. ويوجد عدة أنواع مختلفة من علماء البيانات، مثل أولئك الذين لديهم خلفيات برمجة قوية، أو معرفة إحصائية متعمقة، أو خلفيات واسعة النطاق في قضايا الأعمال. وفي حين تتشارك جميع أنواع علماء البيانات نفس الهدف المتمثل في إنشاء القيم من البيانات، إلا أن بؤرهم تتفاوت في الاختلاف إلى حدٍ ما. ويجيد علماء البيانات الذين يتمتعون بمهارات برمجة قوية تصميم وتنفيذ، خوارزميات وأدوات ونظم جديدة، كما أنهم يسعون إلى الارتقاء بفعالية وكفاءة البرامج / والنظم. وسيقوم تعليم علم البيانات الذي تقدمه أقسام علوم الحاسب بعمل مناسب في هذا الاتجاه. ويجيد علماء البيانات ذوي الخلفيات الإحصائية القوية تفسير/ تحليل البيانات باستخدام التقنيات الإحصائية المتقدمة. ويحاول علماء البيانات ذوي الخلفية التجارية تصويب نماذج الأعمال من البيانات مع التركيز على تحليلات الأعمال. ويركز تعليم علم البيانات، أو تحليلات الأعمال الذي تقدمه العديد من كليات الأعمال على الاتجاهين الأخيرين (الاحصاء والأعمال).



شكل 19 : استخراج المعرفة

وبمقارنته بالأنواع المذكورة أعلاه، يحاول تعليم علم البيانات المقترح القائم على المستخدم إعداد علماء البيانات الذين يمكنهم التفكير من منظور المستخدمين النهائيين في منتجات علم البيانات، أو حل المشكلات التطبيقات بنظرة واسعة الأفق لدورات حياة علم البيانات، من المتطلبات إلى بناء نموذج البيانات، وتقييمها،

ونشرها، ومعالجتها. وتركز معظم مناهج **كليات المعلومات المستقبليه** على مفاهيم مثل: نمذجة المتطلبات، وإدارة واصفات البيانات (الميتاداتا)، ودورات حياة البيانات والبرمجيات، والحوسبة والتصميم المتمركزان حول الإنسان، وتقييم البرمجيات أو النظام، والجوانب الاجتماعية للمعلومات مثل: أخلاقيات البيانات الضخمة وعلم البيانات، وقضايا الأمن والخصوصية الخاصة بالبيانات الضخمة، وتمثيل البيانات وتنظيمها. ومع ذلك، فإن تغطية هذه الموضوعات في مناهج علوم الحاسب، أو كليات إدارة الأعمال ليست واسعة النطاق كما هو عليه الحال في **كليات المعلومات المستقبليه.** ويتميز منهج مدارس المعلومات بمزايا قوية في تعليم علم البيانات المستند إلى المستخدم، من خلال تدريب الطلاب الذين يدركون أهمية نمذجة المتطلبات، ويعرفون أدوار البيانات الوصفية ويستخدمونها، ويصممون ويطورون النظم بما يُراعي تحقيق سهولة الاستخدام التي تركز على الإنسان، والاهتمام بأمن البيانات والخصوصية في جميع مراحل دورة حياة علم البيانات، والتحقق من صحة نتائج التحليل، وأداء سرد القصص المناسبة لأصحاب المصلحة، وإدارة المشروعات مع رؤى دقيقة حول دورات حياة علم البيانات، وحماية البيانات والرؤى التي يتم التوصل إليها، واستخدام البيانات والنتائج بشكلٍ أخلاقي، ومعرفة كيفية وضع الاستراتيجيات اللازمة لأرشفة البيانات وتنظيمها.

ومن أمثلة الأسئلة التي يعالجها علماء البيانات من وجهة نظر المستخدم:

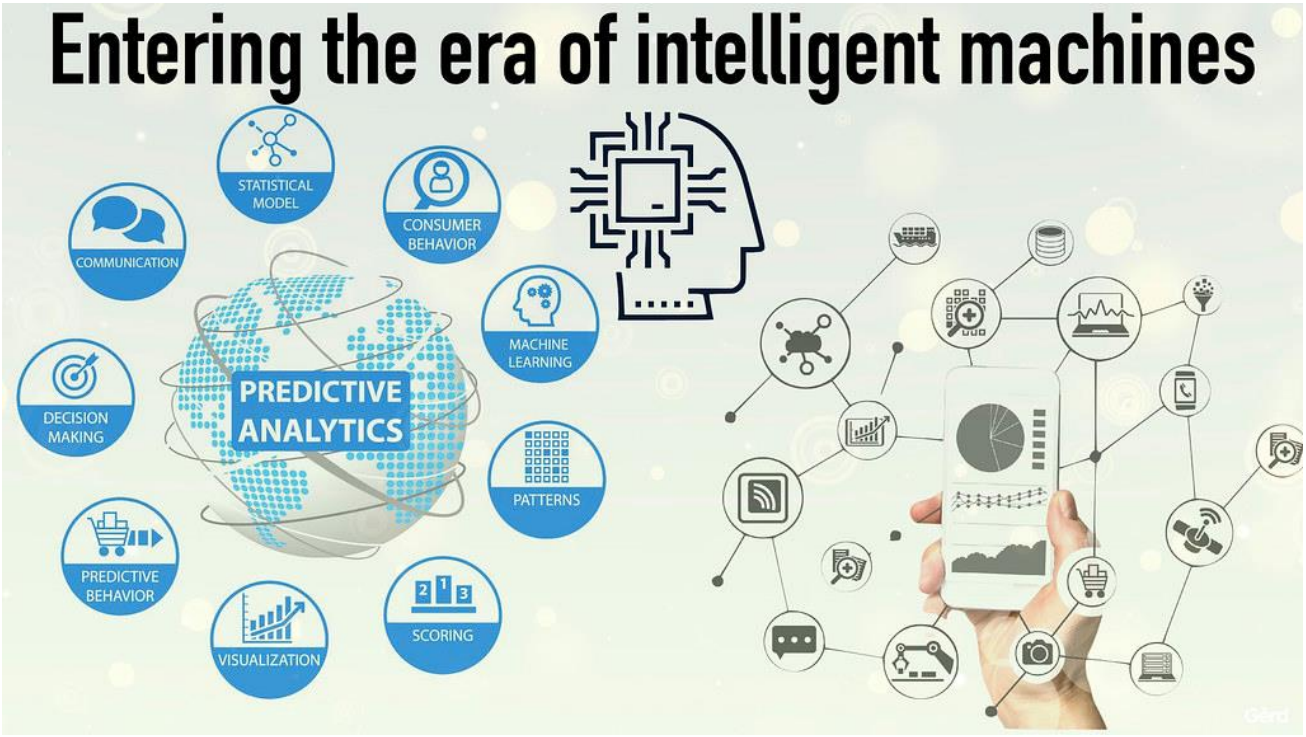
(1) ما هي المشكلات الرئيسية التي يريد المسوقون عبر الإنترنت حلها باستخدام نهج علم البيانات، وما هي الأسئلة المحددة التي يحتاجون إلى طرحها لحل المشكلات ؟

(2) ما هي جوانب منتج علم البيانات المحدد الذي يشعر المستخدمون النهائيون بالراحة وعدم الارتياح معه؟

(3) ما هي أنواع التقارير النهائية / الأفكار / التمثيلات الأكثر فائدة في توصيل المعلومات الأساسية لكل صاحب مصلحة، وكيف يجب تقديمها لهم كقصة؟ لا يمكن معالجة هذه الأسئلة بشكلٍ فعالٍ إلا إذا ركزنا على احتياجات هؤلاء المستخدمين ومتطلباتهم والتحقق من صحتها وفعالية التمثيلات والعروض التقديمية.

واستنادًا إلى هذه الاعتبارات، يجب أن تركز طرق **كليات المعلومات المستقبليه** لتدريس تعليم علم البيانات، وعلى نطاق واسع، على من القائمين فعلياً على حل المشكلات، أو المستخدمين الذين سيستخدمون منتجات علم البيانات. ومن وجهة النظر العملية، وفي حين تتسم جميع برامج علم البيانات بقلة عدد المقررات الدراسية المقدمة خلالها، لن يكون هناك ترف في اتاحة مقررات منفصلة تغطي الموضوعات المذكورة أعلاه. وبالتالي يجب دمج

الأفكار الرئيسية لتلك الموضوعات بشكلٍ استراتيجيٍ في العديد من مقررات علم البيانات الصلبة المختلفة مثل: مقدمة في علم البيانات Introduction to Data Science، وإدارة مشروع علم البيانات Data Science Project Management، والتواصل مع البيانات Communicating with Data، والتمثيل Visualization، ومشروع التخرج الشامل the Capstone Project، وما إلى ذلك.



شكل 20: علم التنبوء بالمستقبل

وتجدر الاشارة هنا إلى التحذير من مشكلة واحدة وهي أن نكون حذرين في عدم تحويل طلابنا إلى اختصاصيين لا يمتلكون خبرة قوية خاصة بهم في مجالات محددة. فعلى الرغم من أن الطلاب يجب أن يكون لديهم فهم واسع حول مختلف القضايا التي سيلتفون حولها أثناء عملهم في مشروعات علم البيانات، فمن الضروري أن يقوم كل طالب بتطوير خبرته الخاصة في بعض الجوانب أو المجالات الخاصة بدورة حياة علم البيانات. ومن أجل تنمية الخبرة الشخصية، يمكن للطلاب العمل مع علماء البيانات الراسخين والانخراط في مشروعات واقعية يقودها موجهون قادرون. ونظرًا لأن مقرر مشروع التخرج يساعد في تحقيق هذا الهدف، فيجب أن يكون مقرراً أساسياً إلزامياً في أي منهج لعلم البيانات.

5-2- تعليم علم البيانات المستند إلى الأداة:

الغرض من تعليم علم البيانات القائم على الأدوات في **كليات المعلومات المستقبليه** هو التأكيد على أهمية الأدوات الآلية واستخدام المكتبات المتاحة. ولا يعني تعليم علم البيانات القائم على الأداة أن الطلاب يحتاجون إلى تعلم الأدوات فقط دون تعلم لغات البرمجة، حيث إنه يجب على طلاب مدارس المعلومات تعلم اللغات التحليلية عالية المستوى جنباً إلى جنب مع أدوات علم البيانات. ويساعد تعلم لغات البرمجة الطلاب على التفكير المنطقي والحاسوبي ، وهي مهارة بالغة الأهمية في أي تعليم يرتبط بتقنية المعلومات. ونوصي هنا بلغات البرمجة النصية عالية المستوى مثل Python، و R ، بدلاً من Java أو C ++. وتتسم لغات Python، و R بسهولة التعلم، وامتلاك مجموعة واسعة من المكتبات والحزم، والتي يمكن مزجها بسهولة مع أطر تحليلات البيانات الكبيرة مثل: Hadoop و Spark وهذا ما قد لا يتوافر Java، و C ++. وتكمن النقطة الأساسية هنا في تطوير التطبيقات باستخدام المكتبات الموجودة أو خدمات الويب، وليس في تنفيذ الخوارزميات، والتي يتم تعلمها في علوم الحاسب. ويجب أن يتعرض الطلاب لمجموعة واسعة من المكتبات والخدمات المتاحة، وأن يتعلموا كيفية استخدامها بشكلٍ فعالٍ، دون الحاجة إلى "إعادة اختراع العجلة". ويتوقع من هذا النهج أن يكسب الطلاب خبرة واسعة في تطوير تطبيقات جديدة مع المكتبات الحالية ومساعدتهم على التركيز على حل المشكلات بدلاً من التكويد منخفض المستوى.

وبالإضافة إلى اللغات عالية المستوى، يجب على الطلاب أن يتعلموا أيضًا واحدة أو أكثر من أدوات علم البيانات الرئيسية. ويوجد ثلاثة أسباب مهمة تدعو للتأكيد على استخدام الأدوات الآلية في تعليم علم البيانات. **أولاً**، سوف يستوعب الطلاب تدفق مشروعات علم البيانات بسرعة أكبر من خلال النقر ودون الحاجة إلى كتابة أية أكواد. وسوف تتعامل الأدوات المؤتمتة مع إدخال البيانات، وبعض عمليات المعالجة المسبقة للبيانات، وتنفيذ الخوارزميات، وتمثيل المخرجات. ويمكن للطلاب باستخدام الأدوات الآلية التركيز على مشكلات الأعمال، وطبيعة بيانات الإدخال، ونقاط القوة والضعف في الخوارزميات التي يتم تنفيذها، وتفسير المخرجات، والتحقق من صحة النتائج. كما أن هذه الأدوات ستسمح لهم بالتقدير السريع لقيمة وأهمية مناهج علم البيانات، ودعم فكرة التعلم عن طريق الممارسة.

ثانيًا، سوف يستخدم معظم الطلاب أدوات أو لغات علم البيانات عالية المستوى في الصناعة لحل مشكلات علم البيانات في العالم الحقيقي في بعض المجالات التطبيقية بمجرد تخرجهم من الكلية. وبالتالي، من المنطقي اتاحة الفرصة لهم لاكتساب خبرة التعامل مع أدوات علم البيانات الآلية الشائعة، وفهم نقاط القوة والقيود الخاصة بها.

ثالثًا، أصبحت الأدوات التلقائية أكثر قوة وتنوعًا، لا سيما مع امتلاكها بعض الميزات مثل: المعالجة المسبقة للبيانات، وبناء النماذج، والتعلم الآلي، والتحقق، والتمثيل، وإنشاء التقارير. ومن الأمثلة على تلك الأدوات: IBM Watson Analytics، PurePredictive، وBigML، وRapidMiner، وDataRobot. وتتضمن نقاط القوة هذه الأدوات منحنى التعلم السطحي نسبيًا، والأتمتة الجزئية أو الكاملة لتنقيب البيانات، أو خوارزميات التعلم الآلي، والكفاءة والفعالية المثبتة. وستصبح هذه الأدوات المؤتمتة تدريجياً أكثر نضجًا وقوةً وموثوقيةً. ومما لا شك فيه أن ميزات الأدوات التي تم تطويرها حديثًا، وتجريبها مع مجموعات البيانات الصغيرة او متوسطة الحجم لأغراض التعلم، وفهم إيجابياتها وسلبياتها، وجمع حالات الاستخدام الناجح للأدوات ستمنح علماء البيانات الموجهين للمستخدم user-oriented data scientists قوة هائلة.

ويجب مراعاة الأدوات الآلية أولاً قبل استخدام اللغات عند تطوير تطبيق معين. وإذا لم تحقق الأدوات درجة عالية من الرضاء، فيجب مراعاة اللغات بعد ذلك. ومن وجهة النظر هذه، يجب ألا تهمل **كليات المعلومات المستقبليه** تدريس الأدوات الموجودة على عدة مستويات (على سبيل المثال، مستوى الخوارزمية، ومستوى المكتبة، ومستوى البرنامج).

ويُعد نقص أعداد علماء البيانات المؤهلين من أبرز المشكلات أكثرها صعوبة في الصناعة والتعليم. ويمكن للنهج القائم على الأدوات أن يساعد في الحد من هذا النقص، عن طريق تدريب المستخدمين ذوي المعرفة المحدودة بعلم البيانات للتصدي لمشكلات علم البيانات. بتمكين المستخدمين بالأدوات الآلية مسمى "**علم بيانات المواطن citizen data science** ". وعلماء بيانات المواطنين هم المستخدمون النهائيون الذين يعمدون إلى حل مشكلات علم البيانات باستخدام الأدوات الآلية دون تشفير. و هذا الاتجاه في قطاعات أخرى مثل، صناعة السيارات، التي تمتلك عدداً من السائقين يفوق بكثير عدد الميكانيكيين.

5-3- تعليم علم البيانات المستند إلى التطبيق:

النقاط الرئيسية لتعليم علم البيانات القائمة على التطبيق في مدارس المعلومات هي ثلاثة أضعاف، حيث تعمل على تطوير: (أ) التعليم القائم على المشروعات، (ب) القدرة على العمل مع خبراء المجال، (ج) اكتساب الخبرة في مجال تطبيق واحد. **أولاً**، يجب أن يشارك الطلاب في مشروعات علم البيانات بأشكالٍ ومستوياتٍ متنوعة. في المقررات السابقة، يجب أن يفهم الطلاب حالات استخدام البيانات الضخمة التي يتم تطبيق تخصصات علم البيانات عليها للتحول الرقمي. كما يجب على الطلاب تجربة العديد من دراسات الحالة الكاملة مع التعرض لكل مرحلة من مراحل دورة حياة علم البيانات، باستخدام أداة آلية في المرحلة المبكرة جدًا من برنامج

الشهادة. ومن شان ذلك أن يعطيهم صورة كبيرة لمشروع علم البيانات بدون تكويد، ويسمح لهم بالتفكير في المجال الذي يريدون العمل فيه بشكلٍ أكبر، ومساعدتهم على اختيار مجال معين يريدون تطوير خبراتهم الخاصة فيه خلال بقية برنامج دراستهم. وبعد ذلك، يجب أن يكون هناك مقرر دراسي أساسي يعمل عليه الطلاب كفريق لمعالجة مشروعات علم البيانات الواقعية، وربما كان ذلك بمشاركة خبير في المجال.

ثانيًا، يجب أن يتعلم الطلاب كيفية العمل مع خبراء المجال وطلاب التخصصات الأخرى كفريق. العمل الجماعي مهم بشكل عام في أي تخصص، ولكن لا يمكن التأكيد على هذا المطلب كثيرًا في تعليم علم البيانات، حيث إن معظم مشكلات علم البيانات معقدة، ومتعددة التخصصات، وتتطلب جهدًا جماعيًا. ويجب على علماء البيانات التواصل بشكلٍ متكررٍ فيما بين بعضهم البعض، لاستيعاب معارفهم الخاصة بالمجال، والعمل على جعل البيانات ومناهج علم البيانات مفهومة من قبل جميع الأطراف، وتمثيل النتائج وتقديمها بطريقة واضحة وخالية من المصطلحات المعقدة. لذلك، تعد القدرة على التعاون مع خبراء المجال، أو الأعضاء الآخرين في الفريق من المهارات الأساسية.

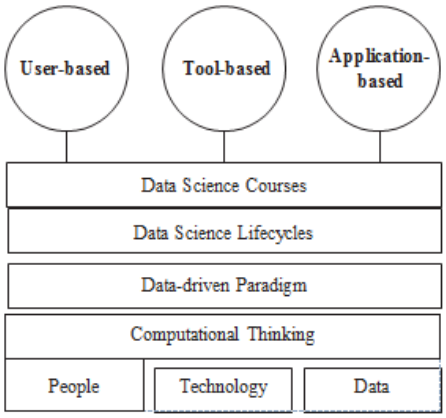
ثالثًا، من المستحسن أن يقوم الطلاب بتطوير قدراتهم أو خبراتهم في مجالٍ واحد. ويتطلب الفهم الكامل المتكامل لمجال واحد وقتًا وجهدًا كبيرين، وبالتالي من المستحيل أن يكون لدى عالم بيانات واحد معرفة متعمقة بالعديد من المجالات. ويجب تشجيع الطلاب على تطوير الخبرة والمعرفة المتعمقة في مجالٍ واحدٍ، لأن معظم مشكلات علم البيانات معقدة وتتطلب خبرة واسعة. ويجب أيضًا تشجيع الطلاب على العمل في المجالات غير التقليدية لإنشاء تطبيقات مبتكرة. على سبيل المثال، يقال أن IBM Watson يمكنه إنشاء وصفات من خلال دمج مكونات طعام متعددة، وتأليف قطعة موسيقية بعد تعلم أسلوب الفنان، وإنتاج اللوحات، وابتكار معاينات لأفلام، وسيناريوهات. و تتجاوز هذه التطبيقات الطرق التجريبية التقليدية مثل: المبيعات، والتسويق، والرعاية الصحية، والسفر، وما إلى ذلك. وسيتيح امتلاك الطلاب الاستعداد للابتكار فرصًا للمساهمة في المجال، بل قد يفتح مجالاتٍ جديدة.

وبشكل عام، يجب أن تركز الاستراتيجيات التعليمية لعلم البيانات على تدريس الرؤى المستندة إلى المستخدم باعتبارها أداة لحل المشكلات في مجالات التطبيق، وأن تتجه إلى تلبية الاحتياجات الخاصة للمستخدمين النهائيين لمنتجات علم البيانات. ويمكن لمدارس المعلومات تعليم كلٍ من: التكويد باستخدام لغة برمجة عالية المستوى، وأدوات علم البيانات المؤتمتة. ويجب أن يوفر المنهج العديد من التجارب القائمة على المشروع حتى يتمكن الطلاب من استكشاف خياراتهم الخاصة والعثور على مجالهم المفضل.

6 . الإطار المقترح لتعليم علم البيانات:

من حيث أطر تعليم علوم البيانات، تضطلع **كليات المعلومات المستقبليه** بتدريس موضوعات متعددة التخصصات مثل: الجوانب البشرية، والتكنولوجية، والمجتمعية للمعلومات. ويتضمن المنهج الحديث النموذجي **لكليات المعلومات المستقبليه** بالفعل العديد من المقررات الأساسية لتعليم عم البيانات مثل: البرمجة، والإحصاءات الأساسية، وإدارة البيانات، وطرق البحث. ويمكن ل**كليات المعلومات** المستقبليه تعزيز تعليم علم البيانات من خلال دمج بعض الموضوعات المتقدمة في المناهج الدراسية، مثل برمجة علم البيانات، والبيانات الضخمة والحوسبة السحابية، وتحليلات البيانات، والتمثيل، والتعلم الآلي، ومشروع التخرج، وغيرها بناءً على تركيز المؤسسة. وبطبيعة الحال، يجب أن يخضع تصميم إعداد منهج علم البيانات لمجالات القوة والتفضيل الخاصة بكل مؤسسة.

في حين يتم تصميم منهج لعلم البيانات، نقترح تنفيذ إطار عملٍ متسقٍ ومتماسك، من خلال تصميم مقررات تغطي الإطار أثناء تدريس الطلاب المبادئ التي تدعم الإطار. ويوضح الشكل 3 إطار العمل المقترح لتعليم علم البيانات في **كليات المعلومات المستقبليه** (DSEF) Data Science Education Framework، حيث تمثل الركائز الثلاثة (الأشخاص، والتقنية، والبيانات) الأسس التي يقوم عليها تعليم علم البيانات. وباعتبارها البنات الأساسية بناء مقررات علم البيانات، فإننا نضع التفكير الحاسوبي، والنموذج القائم على البيانات، ودورة حياة علم البيانات على رأس الركائز الثلاث. فنحن ننظر إلى التفكير الحاسوبي باعتباره طريقة لتنمية مهارات حل المشكلات، والنموذج القائم على البيانات باعتباره مبدأ أساسي لعلم البيانات، ودورة حياة علم البيانات باعتبارها الأساس لإنشاء مقررات علم البيانات. ونقترح أن يتم تدريس مقررات علم البيانات في مدارس المعلومات في ضوء المحاور الثلاث لحوسبة المعلومات – الطرق القائمة على المستخدم ، والأداة ، والتطبيق.



شكل 21: إطار العمل المقترح لتعليم علم البيانات في كليات المعلومات المستقبليه.

6-1- الركائز الثلاثة لعلم البيانات:

مثلما ركزت مدارس علم المعلومات وخلفائها الجدد، وهم **كليات المعلومات المستقبليه** ، على الأشخاص والتقنية والمعلومات، فنحن نوصي أيضًا بضرورة النظر إلى الأشخاص والتقنية والبيانات باعتبارها خط أساس لتعليم علم البيانات في **كليات المعلومات المستقبليه**ويتمثل الاختلاف الوحيد في تعليم علم البيانات في التركيز على البيانات بدلاً من المعلومات، من أجل التأكيد على النموذج القائم على البيانات الذي نعمل في سياقه حاليًا. وبناءً على هذه الركائز الثلاث لعلم البيانات، نقترح فرض التفكير الحاسوبي كإطار أساسي لحل المشكلات. ويجب بعد ذلك تصميم مقررات علم البيانات بحيث تغطي على نطاق واسع القضايا والتقنيات الموجودة في دورة حياة علم البيانات. ويجب أن تستوعب الطرق التي يتم بها تدريس مقررات علم البيانات مبادئ حوسبة المعلومات في أشكالٍ تعتمد على المستخدم، والأداة، والتطبيق، كما هو مذكور بعاليه.

6-2- إطار حل المشكلات: التفكير الحاسوبي:

يحتاج الطلاب، بعد أن يتم إمدادهم بالمعرفة الأساسية اللازمة، إلى معرفة كيفية حل مشكلات البيانات الضخمة باستخدام التفكير الحاسوبي. ويعرّف **التفكير الحاسوبي بأنه "طريقة لحل المشكلات، وتصميم النظم، وفهم السلوك البشري الذي يعتمد على المفاهيم الأساسية لعلوم الحاسب.** إنه يمثل مجموعة من الاتجاهات والمهارات القابلة للتطبيق على المستوى العالمي، والتي يتوق الجميع- وليس فقط علماء الحاسب- إلى تعلمها واستخدامها.. ويشير التفكير الحاسوبي إلى مجموعة من عمليات التفكير اللازمة لحل المشكلات باستخدام أجهزة الحاسب، والتي تستخدم لتطوير جميع منتجات البرامج. ونظرًا لوجود العديد من منتجات البرامج التي سيتم تطبيقها من قبل المستخدمين النهائيين، فإنها تعتبر مهارة أساسية للجميع، خاصة وأن البرمجيات أصبحت واسعة الانتشار وأكثر أهمية من الأجهزة. ويُعرف التفكير الحاسوبي بالفكر التحليلي الذي يشارك بطرقٍ عديدة في التفكير الرياضي (حل المشكلات)، والتفكير الهندسي (تصميم النظم) والتفكير العلمي (فهم السلوك البشري). ومع المفاهيم الأساسية مثل: التجريد والتحليل والأنماط،و **يمكن إعادة صياغة أية مشكلة كبيرة ومعقدة وصعبة إلى مشكلات نعرفها، ويمكن حل هذه المشكلات بشكلٍ أكبر باستخدام الخوارزميات والبيانات. إن "التفكير الحاسوبي يؤثر على البحث في جميع التخصصات تقريبًا، في كل من العلوم والانسانيات على حدٍ سواء" إن وجود العديد من التخصصات التي تأثرت بوضوح بالتفكير الحاسوبي مثل: الإحصاء، والبيولوجيا، والاقتصاد، والكيمياء، والفيزياء. ونظرًا لأن التكويد هو أفضل طريقة لتعلم التفكير الحاسوبي، يجب على الطلاب تعلم الترميز كوسيلة لحل المشكلات بشكلٍ منهجيٍ والتفكير بشكلٍ منطقي.**

6-3- النموذج المستند إلى البيانات:

المبدأ الأساسي هو النموذج القائم على البيانات the data–driven paradigm (DDP). من المعتقد أن البيانات هي أحد الأصول، وأن البيانات تحكي القصة، بما يؤدي إلى صنع القرار القائم على البيانات. وتتمتع **كليات المعلومات المستقبليه** بموقع فريد لتحفيز الطلاب على فهم النموذج القائم على البيانات. ويجب على الطلاب أن يحاولوا التعرف على ما يمكن تقودنا إليه البيانات الضخمة، وكيف يتم استخدام البيانات في أساليب علم البيانات، وما هي آثار النموذج الذي يعتمد على البيانات على العلوم، والهندسة، والصناعة، والمجتمع. ولقد تمت مناقشة التغييرات التي أدخلتها البيانات الضخمة على المجتمع على نطاق واسع .

وتعد فكرة تحويل البيانات Datatification إحدى النقاط الهامة. وينطوي تحويل البيانات على تصميم بيانات يمكن معالجتها آليًا عن طريق رقمنة الأفكار، أو سلوكيات البشر، أو الآلات، أو الأشياء. يمكن أن تشمل بعض الأنشطة المهمة هنا القياس الكمي للبيانات الدقيقة وقياسها ، وتحويل الظاهرة إلى بيانات كمية للجدولة والتحليل. يمكن أن يكون أحد الأمثلة على ذلك بيانات البيانات عن الحياة الشخصية ، أي سلوكيات مستخدم بيانات البيانات مثل الموسيقى التي يستمعون إليها وكم مرة ومتى ، وعدد مرات تشغيل المستخدم للسخانات ، وعدد المرات التي يذهب فيها الأشخاص لشراء الطعام ومتى ، وما إلى ذلك. إنشاء بيانات جديدة مجموعة عبر البيانات ستكون أساسا للتطبيقات المبتكرة في العديد من الصناعات والتخصصات.

ويمكن أن تشمل بعض الأنشطة المهمة هنا القياس الكمي للبيانات الدقيقة وعايرتها، **وتحويل ظاهرة ما إلى بيانات كمية من أجل الجدولة والتحليل.** ويمكن أن يكون أحد الأمثلة على ذلك تحويل البيانات عن الحياة الشخصية ، أي استجماع البيانات عن سلوكيات مستخدم بيانات مثل: ما هي الموسيقى التي يستمعون إليها وعدد مرات الاستماع وفي أي الأوقات، وعدد مرات تشغيل المستخدم للسخانات، وعدد المرات التي يذهب فيها الأشخاص لشراء الطعام وفي أي الأوقات، .. إلخ. وسوف يمثل إنشاء مجموعة بيانات جديدة عن طريق تحويل البيانات أساسًا للتطبيقات المبتكرة في العديد من الصناعات والتخصصات.

ويجب على الطلاب أن يفهموا بوضوح كيف يتيح النموذج القائم على البيانات فرصًا جديدة لم تكن ممكنة في الماضي. وتعد جوجل Google أبرز مثال على شركة طبقت البيانات كنموذج خدمة. وإلى جانب الحفاظ على مكانتها كمحرك البحث الأول ، **قامت جوجل بتحويل كلمات البحث الأساسية الخاصة بنا إلى ما يسمى اتجاهات جوجل، واستخدمت البيانات المجمعة لترجمة لغة إلى أخرى. وتمكننا البيانات الضخمة من طرح أسئلة تجارية جديدة واستنباط قرارات داعمة.** وفي كل وقتٍ يظهر فيه مصدر بياناتٍ جديد، يتحتم علينا التفكير في كيفية استخدام هذا المصدر وتحديد البيانات التي يمكن تدمج معه لخلق نوع من التكاملية. على سبيل المثال، يتم استخدام بيانات الشبكات الاجتماعية، التي تتضمن سلوكيات المستخدمين على نطاق واسع، لأغراض عديدة

ومختلفة منها: **التسويق الاجتماعي، ومبيعات الهواتف المحمولة، والتنقيب في الآراء، وتحليل الانتخابات، وتحديد المستخدمين الأكثر تأثيرًا**، وما إلى ذلك. **ويتم تحليل بيانات نقاط البيع التقليدية مع بيانات الطقس لمعرفة آثار الطقس على المبيعات. ويمكن دمج بيانات تويتر Twitter مع بيانات الطقس وبيانات الرعاية الصحية الإلكترونية للتنبؤ بعدد مرضى الربو الذين يترددون على غرف الطوارئ**، وما إلى ذلك. فعندما يتم دمج مصدر البيانات مع مصدر بيانات آخر، يمكن الخروج برؤى جديدة.

ويؤدي النموذج القائم على البيانات أيضًا إلى ايجاد الاستكشاف العلمي قائم على البيانات، والاقتصاد القائم على البيانات. وفي حين لا تعد ممارسة الطب القائم على الأدلة في المجالات الطبية أمراً جديداً، إذ يتم ممارسته منذ أمد بعيد، إلا أن إدارة علم البيانات تفرض الاستثمار المكثف لهذا المنحى. وأصبحت تطبيقات تقنيات علم البيانات على البيانات العلمية الضخمة الآن نموذجًا جديدًا في كلٍ منك علم الفلك، والفيزياء، وعلم الأحياء. ويعتبر علم البيانات أحد المجالات سريعة التغير، التي تتطلب المتابعة المستمرة لآخر تطورات البحث العلمي. و يساعد هذا على تجهيز مشروعات علم البيانات بأفضل التقنيات المتاحة والفعالة في ضوء التجارب السابقة والاتجاهات المتطورة.

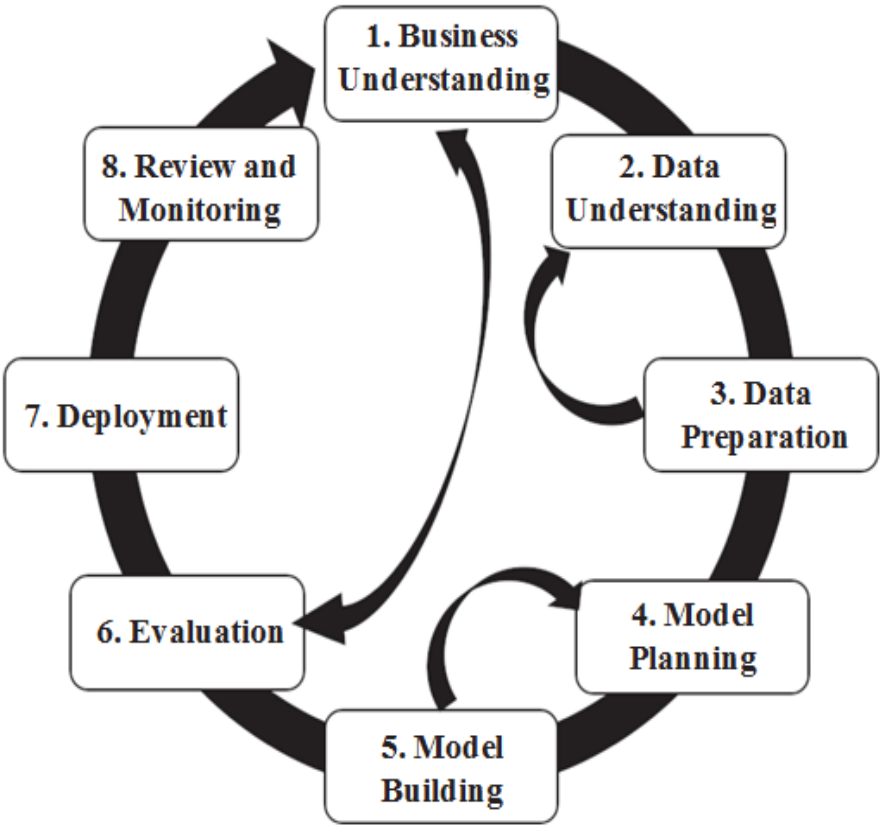
وبالإضافة إلى الاكتشافات العلمية التي تعتمد على البيانات، فإن فهم الأسواق القائمة على البيانات، والخدمات التي تعتمد على البيانات، والاقتصاد القائم على البيانات ضروري أيضًا لتعلم الطلاب. وتقوم شركات تقنيات المعلومات الرئيسية مثل: Google، وFacebook، وIBM، وAmazon بالعمل على زيادة أصول بياناتها بسرعة، من خلال الاستحواذ على الشركات التي قامت بتجميع البيانات في الأسواق والعلاقات المهنية (مثل: LinkedIn التي حصلت عليها Microsoft)، والطقس، والأخبار، والموضوعات الأخرى ، أثناء جمع وتحليل كميات هائلة من البيانات، للعثور على فرص جديدة للتطبيقات المبتكرة. وفي الوقت الحاضر، يمنح العديد من العملاء قيمة للخدمات التي تقدمها المنتجات أكثر من المنتجات نفسها، ويساعد علم البيانات على إنشاء خدمات تعتمد على البيانات عالية الجودة. **وتعد الأنواع الجديدة من الشركات مثل: Uber و Airbnb أمثلة جيدة على الاقتصاد القائم على البيانات، والذي يستخدم البيانات لبيع خدمات الأعمال، مما تسبب في الاضرار بأعمال سيارات الأجرة والفنادق بشكلٍ غير مسبوق.**

ويجب أن يتعلم الطلاب أيضًا العديد من حالات استخدام البيانات الضخمة، بما في ذلك كيفية تطبيق النموذج القائم على البيانات في حالات الاستخدام هذه. وسوف يساعد استيعاب الطلاب للحالات الموجودة على إنشاء حالة جديدة لاستخدام البيانات الضخمة وأفكار البيانات. **وعندما يتقن الطلاب طرق تحديد المشكلات الكبيرة في البيانات الضخمة من خلال الجمع بين مصادر متعددة للبيانات الضخمة، وتطبيق التفكير المنهجي**

لايجاد معارف، أو منتجات، أو خدمات جديدة، فسيكون بإمكانهم الانضمام إلى المتخصصين مجال في البيانات الضخمة.

6-4- دورات حياة علم البيانات:

يمكن تلخيص الخطوات المحددة لحل مشكلات البيانات كدورة حياة علم البيانات، أو دورة حياة تحليلات البيانات. اقترحنا دورة حياة تحليلات البيانات في دراستنا السابقة (انظر الشكل 6 في Song & Zhu, (2016)، والتي نسميها في هذه الورقة بدورة حياة علم البيانات. وكما هو مبين في الشكل 4، تتضمن الخطوات الثماني الهامة فهم الأعمال، وفهم البيانات، وإعداد البيانات، وتخطيط النموذج، وبناء النموذج، والتقييم، والنشر، والمراجعة والرصد (الجدول 1).



شكل 22: دورة حياة علم البيانات.

جدول (1): الخطوات الثمانية لدورة حياة علم البيانات.

م	الخطوة	الخطوة الفرعية
1-	**فهم الأعمال Business understanding**	ما هو السؤال الذي يجب حله، وما هي المقاييس التي يجب تقييمها؟ وتصميم الفروض، وتقييم المصادر (الأشخاص، والبياناتن والأدوات).
2-	**فهم البيانات Data understanding**	تحديد مصادر البيانات، وإعادة استخدام البيانات وخطة التكامل، وتحويل البيانات، واتخاذ القرار بشأن الأدوات.
3-	**إعداد البيانات Data preparation**	الحصول على البيانات، والقيام بتنميط البيانات وتنظيفها وتحويلها، واستكشاف البيانات والتحقق من الجودة.
4-	**تخطيط النموذج Model planning**	تحديد الأساليب والتقنيات وسير العمل، وتحديد المتغيرات الرئيسية والعلاقة الارتباطية بينها.
5-	**بناء النموذج Model building**	بناء النماذج ؛ إجراء التحليل والتكرار.
6-	**التقييم Evaluation**	إجراء تقييم مقابل المقاييس، وايصال النتائج والتوصيات.
7-	**النشر Deployment**	دمج إجراءات التحليلات في لوحات تحكم الإدارة، ونظم التشغيل.
8-	**المراجعة والرصد Review and monitoring**	مراقبة الأداء، و تحديد الأجزاء التي تحتاج إلى تحسين.

ويمكن **لكليات المعلومات المستقبليه** تضمين جميع خطوات دورة حياة علم البيانات في منهج علم البيانات والمقررات الهيكلية المستندة إليها. ويمكن أن يكون لمقرر مشروع التخرج لعلم البيانات أن يكون منصة جيدة ، تتيح للطلاب ممارسة جميع خطوات دورة الحياة.

6- 5- موضوعات علم البيانات الخاصة بمدارس المعلومات:

6- 5- 1- إدارة مشروع علم البيانات:

تعتبر إدارة مشروع علم البيانات واحدة من أهم مهارات علم البيانات، حيث ترتبط ارتباطًا وثيقًا بدورة حياة علم البيانات. وتتطلب إدارة البيانات الضخمة ومشروعات علم البيانات بنجاح، الإلمام الشامل بتقنيات البيانات الضخمة، ودورات حياة علم البيانات، وتقنيات علم البيانات، ومهارات الاتصال، ومجالات التطبيق. وفضلاً عن ذلك يجب أن يكون لدى المرء القدرة على تحديد مشكلات الأعمال، وطرح أسئلة محددة حول مشكلات الأعمال، وتحديد مصادر البيانات المناسبة، واختيار أدوات علم البيانات والمنصات التحليلية المناسبة، وتقييم النتائج، والتواصل مع أصحاب المصلحة. وقد ينظر مديرو مشروعات علم البيانات أيضًا في جمع البيانات حول الأشياء لإنشاء آثار أقدام رقمية. ويُطلق على كبير الموظفين الذي يقود ويدير مشروع البيانات الضخمة وعلم البيانات رئيس مسؤولي البيانات Chief Data Officer (CDO). ويكفل بناء منهج دراسي لعلم البيانات على رأس دورة حياة علم البيانات، ، فإن طلاب **كليات المعلومات المستقبليه** سوف تمتلك إمكانات غير محدودة لتعليم المهارات والخصائص المطلوب توافرها في رئيس مسؤولي البيانات.

6- 5- 2- معالجة البيانات الضخمة:

تنطوي البيانات الضخمة على بيانات منظمة، وغير منظمة. وبالتالي، يُعد تنظيم البيانات في البيانات الضخمة أكثر تعقيدًا من الحالات التي يقتصر التعامل فيها مع البيانات المنظمة فقط. ولا تتجاوز النظرة الضيقة لمعالجة البيانات عمليات أرشفة البيانات، و التخزين المستديم لها بما ييسر استرجاعها في المستقبل. وفي سياق البيانات الضخمة، يمكن توسيع نطاق معالجة البيانات ليشمل عمليات جمع البيانات من مصادر مختلفة، ودمجها، وتنظيمها، والتعليق عليها، ونشرها من أجل تتبع المصدر الأصلي للبيانات. وترتبط هذه الأنشطة ارتباطًا وثيقًا بالحفاظ على جودة البيانات وإدارة الميتاداتا. وفي حين تعد جودة البيانات مطلبًا مهمًا لانجاز مشروعات علم البيانات بنجاح، تحظى إدارة الميتاداتا بأهمية خاصة لمعالجة البيانات. وتتضمن التحديات التي تواجه معالجة البيانات الضخمة ضرورة تحديد:

(1) مقدار البيانات التي سيتم تخزينها،

(2) كم سيكلف البيانات المختزنة،

(3) كيف سيتم تأمين البيانات،

(4) ما مدة الحفاظ علي البيانات،

(5) كيف تتم عملية المعالجة بشكلٍ تلقائي،

(6) كيفية إزالة التكرار في البيانات المنظمة،

(7) كيفية معالجة البرامج والتطبيقات،

(8) ما هو النطاق اللازم لمعالجة البيانات،

(9) كيفية التجميع التلقائي للميتادتا الخاصة بالبيانات المنتظمة في مجموعات بيانات مختلفة،

(10) كيفية التعليق التفسيري للبيانات المنظمة لتيسير استخراجها وإدارتها. ولسوف يكون لمواجهة هذه التحديات أثرٌ كبيرٌ على البيانات الضخمة والتطورات البيئية لعلم البيانات.

7- الخلاصة:

لقد أصبح علم البيانات أحد أكثر المجالات تحديًا في الوقت الراهن، ، ما ساهم في ايجاد طلبٍ كبيرٍ على علماء البيانات الموهوبين. وعملت العديد من الكليات والأقسام على تلبية هذا الطلب بسرعة من خلال تدريس مقررات في علم البيانات. وكانت **كليات المعلومات المستقبليه** رائدة على المستوى المؤسسي من خلال تطوير مجتمعٍ يطور تعليم علم البيانات عبر العديد من التخصصات. وتساعد أوجه التشابه بين تخصصات المعلومات وعلم البيانات، على امتلاك مدارس المعلومات لنقاط قوة في تعليم علم البيانات. ونعتقد بأن مدارس المعلومات ستكون قادرة على تعليم العديد من علماء البيانات الناجحين من خلال الاستفادة الفعالة من النقاط القوة المشار إليها.

وفي هذه الورقة، تبدو إحتياجات تعليم علم البيانات إلى مواكبة التطورات الحديثة في البيانات الضخمة، والثورة الصناعية.الرابعة. ويتعين على **كليات المعلومات المستقبليه** ، لمواكبة أحدث الاتجاهات، تمكين الطلاب باستخدام حوسبة المعلومات من خلال تعليمهم إنشاء القيم، والمعلومات، والمعرفة. وقدمت الورقة إطار عمل لتعليم علم البيانات متعدد الطبقات مع ثلاث لبنات بناء أساسية تشكل أساس علم البيانات: التفكير الحاسوبي، والنموذج القائم على البيانات، ودورة حياة علم البيانات. ويمكن القول بان **كليات المعلومات المستقبليه** يمكنها القيام بتدريس مقررات علم البيانات بطرق تستند إلى المستخدم، وإلى الأدوات، وإلى التطبيقات. وفي حين يوجد العديد من التحديات التي يتعين على مدارس المعلومات التغلب عليها لتعليم علم البيانات بشكلٍ فعال، يمثل كل تحد أيضًا فرصة للابتكار. ونأمل أن يساعد إطار تعليم علم البيانات الذي تمت مناقشته في هذه المقالة الزملاء في

كليات المعلومات المستقبليه على تصميم مناهج جديدة لعلم البيانات؛ تلبي احتياجات المستخدمين في الصناعة والمجتمع.

أهم الدورات التدريبية الجديدة لعلوم البيانات:

- **OpenAI Assistants with OpenAI Python API**: Learn to use the new OpenAI Assistants API, allowing GPT models to run code, read your files, and call functions!
- **Computer Vision Web Development: YOLOv8 and TensorFlow.js**: OpenCV.js, TensorFlow.js, YOLOv8, Object Detection, Custom Object Detection, Computer Vision and Web Integration
- **Python & Google Trends: Best for Data Science and Marketing**: Python, Google Trends, Data Science, Analytics, Marketing, Business
- **Unlocking the Power of ArcPy: GIS Automation in ArcGIS Pro**: ArcPy Magic in ArcGIS Pro
- **Introduction to Point cloud processing with Python**: Point Clouds, Python, Lidar, Photogrammetry, 3D modelling
- **TensorFlow Developer Certificate Exam Practice Tests 2024**: Fast-Track to TensorFlow Certification: Practical Tests and Expert Tips for Success
- **Mastering Data Wrangling with PySpark in Databricks**: From Beginner to Pro: Learn Key Data Processing Skills and Machine Learning with PySpark in Databricks

المراجع

1. Il-Yeol Song & Yongjun Zhu (2017). Big Data and Data Science: Opportunities and Challenges of iSchools. Vol. 2 No. 3, 2017. pp 1–18. DOI: 10.1515/jdis-2017-0011
2. Baweja, B., Donovan, P., Haefele, M., Siddiqi, L., & Smiles, S. (2016). Extreme automation and connectivity: The global, regional, and investment implications of the Fourth Industrial Revolution. UBS White Paper for the World Economic Forum Annual Meeting 2016. Retrieved on October, 1, 2016, from https://www.ubs.com/global/en/about_ubs/follow_ubs/highlights/davos-2016.html.

3. Bundy, A. (2007). Computational thinking is pervasive. Journal of Scientific and Practical Computing, 1(2), 67–69.
4. Davenport, T.H., & Patil, D.J. (2012). Data scientist: The sexiest job of the 21st century. Harvard Business Review, 90(10), 70–76.
5. Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64–73. Gartner, Inc. (2016). Organizing for big data through better process and governance. Retrieved on September 15, 2016, from https://www.gartner.com/doc/3002918?ref=SiteSearch&sthkw= citizen%20data%20scientist&fnl=search&srcId=1-3478922254.
6. Kagermann, H., Helbig, J., Hellinger, A., & Wahlster, W. (2013). Recommendations for imple- menting the strategic initiative Industrie 4.0: Securing the future of the German manufacturing industry; final report of the Industrie 4.0 Working Group. Retrieved on September 15, 2016, from http://www.manufacturing-policy.eng.cam.ac.uk/documents-folder/policies/germany- recommendations-for-implementing-the-strategic-initiative-industrie-4-0-bmbf-aquatic/ view.
7. Lasi, H., Fettke, P., Kemper, H., Feld, T., & Hoffmann, M. (2014). Industry 4.0. Business & Information Systems Engineering, 6(4), 239–242.
8. Mayer-Schonberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Boston, MA: Houghton Mifflin Harcourt.
9. Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51–59.
10. Schwab, K. (2016). The fourth industrial revolution: What it means, how to respond. World Economic Forum. Retrieved on October 20, 2016, from https://www.weforum.org/ agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond.
11. Stanton, J. (2012). An introduction to data science. Retrieved on September 15,2016, from http://ischool.syr.edu/media/documents/2012/3/datasciencebook1_1.pdf.
12. Song, I.-Y., & Zhu, Y. (2016). Big data and data science: What should we teach? Expert Systems, 33(4), 364–373.
13. Storey, V., & Song, I.-Y. (2017). Big data technologies and management: What conceptual model- ing can do? Data & Knowledge Engineering, 108, 50–67. Retrieved on October, 1, 2016, from https://www.ubs.com/global/en/about_ubs/follow_ubs/ highlights/davos-2016.html.
14. Bundy, A. (2007). Computational thinking is pervasive. Journal of Scientific and Practical Computing, 1(2), 67–69.
15. Davenport, T.H., & Patil, D.J. (2012). Data scientist: The sexiest job of the 21st century. Harvard Business Review, 90(10), 70–76.
16. Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64–73. Gartner, Inc. (2016). Organizing for big data through better process and governance. Retrieved on September 15, 2016, from https://www.gartner.com/doc/3002918?ref=SiteSearch&sthkw= citizen%20data%20scientist&fnl=search&srcId=1-3478922254.
17. Kagermann, H., Helbig, J., Hellinger, A., & Wahlster, W. (2013). Recommendations for imple- menting the strategic initiative Industrie 4.0: Securing the future of the German manufacturing industry; final report of the Industrie 4.0 Working Group. Retrieved on September 15, 2016, from http://www.manufacturing-policy.eng.cam.ac.uk/documents-folder/policies/germany- recommendations-for-implementing-the-strategic-initiative-industrie-4-0-bmbf-aquatic/ view.

18. Lasi, H., Fettke, P., Kemper, H., Feld, T., & Hoffmann, M. (2014). Industry 4.0. Business & Information Systems Engineering, 6(4), 239–242.
19. Mayer-Schonberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Boston, MA: Houghton Mifflin Harcourt.
20. Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. Big Data, 1(1), 51–59.
21. Schwab, K. (2016). The fourth industrial revolution: What it means, how to respond. World Economic Forum. Retrieved on October 20, 2016, from https://www.weforum.org/ agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond.
22. Stanton, J. (2012). An introduction to data science. Retrieved on September 15, 2016, from http://ischool.syr.edu/media/documents/2012/3/datasciencebook1_1.pdf.
23. Song, I.-Y., & Zhu, Y. (2016). Big data and data science: What should we teach? Expert Systems, 33(4), 364–373.
24. Storey, V., & Song, I.-Y. (2017). Big data technologies and management: What conceptual modeling can do? Data & Knowledge Engineering, 108, 50–67.
25. Wing, J.M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35.
26. Wing, J.M. (2008). Computational thinking and thinking about computing. Philosophical Transac- tions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366(1881), 3717–3725.
27. Zhu, Y., Yan, E., & Song, M. (2016). Understanding the evolving academic landscape of library and information science through faculty hiring data. Scientometrics, 108(3), 1461–1478.
28. Wing, J.M. (2008). Computational thinking and thinking about computing. Philosophical Transac- tions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366(1881), 3717–3725.
29. Zhu, Y., Yan, E., & Song, M. (2016). Understanding the evolving academic landscape of library and information science through faculty hiring data. Scientometrics, 108(3), 1461–1478.