مفهوم علم البيانات

أ. د. محمد فتحى عبدالهادى
أستاذ علم المعلومات
كلية الآداب – جامعة القاهرة

المستخلص :

علم البيانات من العلوم الحديثة، وقد نمى وازدهر بسرعة كبيرة فى السنوات القليلة الماضية نظرا لارتباطه بمجالات علمية عديدة من ناحية وانتفاعه من التطورات المتلاحقة فى مجال تكنولوجيا المعلومات والاتصالات من ناحية أخرى. تسعى هذه الدراسة الى التعريف بعلم البيانات من منظورات متعددة من أجل استخلاص تعريف يمثل وجهة نظر علم المكتبات و المعلومات ، وهو أن علم البيانات هو علم بينى متعدد الارتباطات يستخدم التكنولوجيات الحديثة من طرق و معالجات و خوارزميات من أجل التعامل مع البيانات المهيكلة و غير المهيكلة و خاصة البيانات الضخمة بهدف استخراج أو استخلاص معرفة و رؤى لخدمة اتخاذ القرارات السليمة و المناسبة والبحث العلمى، وتتناول الدراسة نشأة علم البيانات وتطوره عبر نحو خمسين عاما حيث تم تقديم المصطلح لأول مرة عام 1974 ، فضلا عن أهميته وقيمته كمجال من المجالات الحيوية فى الوقت الحاضر ، وتُبرز الدراسة موضوعات علم البيانات التى تغطى دورة حياة البيانات و ما يرتبط بها وأبرزها عمليات علم البيانات و أدواته و برمجياته و تقنياته و مؤسساته ، و تهتم الدراسة بعلم مكتبات البيانات ، وعلاقة علم البيانات بالمجالات الأخرى وخاصة الاحصاء و علم الحاسب و علم المعلومات .

الكلمات المفتاحية :

علم البيانات؛ البيانات؛ علم المعلومات؛ مجالات اهتمام علم البيانات؛ دورة حياة البيانات.

تاريخ استلام البحث: 2022/10/4 تاريخ إجازة البحث: 2022/10/5

1. تعريف علم البيانات

توجد تعريفات عديدة لعلم البيانات data science من منظورات مختلفة مثل علم الحاسب والإحصاء وعلم المعلومات.

من التعريفات العامة:

- مجال متعدد الارتباطات يستخدم الطرق العلمية والعمليات والخوارزميات والنظم لاستخراج المعرفة والبصائر أو الرؤى insights من البيانات المهيكلة وغير المهيكلة، ويطبق المعرفة من البيانات عبر مدى عريض من ميادين التطبيق (Data science - Wikipedia).
- ميدان دراسة يتعلق بكميات وفيرة من البيانات ويستخدم أدوات وتقنيات حديثة لايجاد أو لاكتشاف أنماط غير مرئية ويشتق معلومات ذات معنى ويعمل قرارات الأعمال، وهو يستخدم خوارزميات تعلم الآلة المعقدة لبناء نماذج تنبؤية ، و البيانات المستخدمة للتحليل يمكن أن تأتى من مصادر مختلفة كثيرة و تُقدم أو تُعرض فى صيغ متنوعة (What is data science, 2022)
- دراسة البيانات، من أين جاءت، وماذا تمثل، والطرق التى تمكّن من تحويلها إلى مدخلات وموارد قيمة لإنشاء استراتيجيات الأعمال وتكنولوجيا المعلومات (What is data science & how does it works, 2022).
- مجال عريض يشير إلى العمليات والنظريات والمفاهيم والأدوات والتكنولوجيات، التى تمكّن من عرض وتحليل واستخراج المعرفة والمعلومات القيمة من البيانات الخام، وهو يعمل على مساعدة الأفراد والمؤسسات على اتخاذ قرارات أفضل من البيانات المختزنة والمستهلكة والمدارة (What is data science?, 2021)

ومن التعريفات المعجمية:

- استخدام الطرق العلمية للحصول على معلومات نافعة من بيانات الحاسب وخاصة الكميات الكبيرة من البيانات (Data science-Cambridge English Dictionary)
- دراسة البيانات، و يتضمن تطوير الطرق لتسجيل واختزان وتحليل البيانات من أجل استخراج المعلومات النافعة. والغرض من علم البيانات هو الحصول على بصائر

ومعرفة من أى نوع من البيانات سواء مهيكلة أو غير مهيكلة (Data science definition).

- مجال متعدد التخصصات يركز على استخراج معلومات مفيدة ورؤى من البيانات عن طريق عملية الاكتشاف أو اختبار الفرضيات (معجم البيانات والذكاء الاصطناعى، 2022، ص59).

ومن التعريفات التى قدمتها بعض المؤسسات :

- دراسة البيانات لاستخراج بصائر أو رؤى ذات معنى للأعمال. إنه نهج متعدد التخصصات يضم المبادئ والممارسات من مجالات الرياضيات والإحصاء والذكاء الاصطناعى وهندسة الحاسب لتحليل كميات كبيرة من البيانات. ويساعد هذا التحليل علماء البيانات على الإجابة عن أسئلة مثل: ماذا حدث، لماذا حدث، ماذا سيحدث وما الذى يمكن عمله مع النتائج (What is data science- Amazon AWS).
- نهج متعدد التخصصات لاستخلاص أو لاستخراج بصائر فعلية من كميات كبيرة ومتزايدة من البيانات المجمعة والمنشأة بواسطة المؤسسات، ويضم علم البيانات إعداد البيانات للتحليل والمعالجة وأداء تحليل البيانات المتقدم وعرض النتائج لكشف الأنماط وتمكين أصحاب المصلحة لاستخراج الاستنتاجات (What is data science?- Ibm).
- علم البيانات يجمع بين تكنولوجيات وتقنيات ونظريات شتى من مجالات متنوعة معظمها متعلق بعلم الحاسب والإحصاء للحصول على المعرفة الفعلية من البيانات، وهو يعمل على استخراج معرفة فعلية مباشرة من البيانات خلال عملية الاكتشاف أو صياغة فرض واختباره (NIST,2015) .

وقد ساهم عديد من الباحثين بتعريفات لعلم البيانات تمثل وجهات نظر متعددة وبعضهم من قطاع المكتبات والمعلومات. ومن نماذج هذه التعريفات:

- مجموعة من المبادئ الأساسية التى تدعم و توجه استخراج أو استخلاص المعلومات و المعرفة من البيانات (Provost; Fawcett, 2013,p.52).

- المقدرة على استخراج المعرفة والرؤى من مجموعات بيانات كبيرة ومعقدة (Patil, 2015).
- مجال بينى يجمع معا الإحصاء وعلم الحاسب وعلم المعلومات ويعتمد بكثافة على النماذج الاحتمالية وتنقيب البيانات وتعلم الآلة للمساعدة على فهم واستخدام الكم الهائل من البيانات التى تنشأ اليوم (Cervone, 2016, p.8)
- أحد المجالات الناشئة متعددة الارتباطات التى تهتم بتحديد واستخلاص الأنماط القيمة من البيانات الضخمة، وذلك بواسطة تحويل البيانات إلى معلومات ثم إلى معرفة من خلال تحليل البيانات والتنقيب فيها (وانج، 2019، ص314).
- علم البيانات هو عن توليد رؤى قيمة من البيانات من أجل عمل قرارات قائمة على المعرفة (Du; Khan, 2020, p.1)
- مصطلح علم البيانات فى صيغته الحالية يشتمل على مدى متنوع من المداخل المقادة بالبيانات حاسوبيا لتحليلات الأعمال و البحوث. وفى الوسط الأكاديمى فإن هذا الفضاء البينى التخصصات يتضمن مدى عريض من المنهجيات تشمل تعلم الآلة وتحليل وسائل التواصل الاجتماعى، وتحليل النص وتحليلات الويب وما إلى ذلك (Herndon, 2022, p.xix-xx) .

ومن المساهمات العربية فى التعريف بعلم البيانات:

"فن تحويل البيانات إلى أفعال تتحقق من خلال إنشاء منتجات بيانات تقدم معلومات فعلية بدون تعريض متخذى القرارات إلى بيانات أو تحليلات محددة، واستخلاص معلومات فعلية وفورية مفيدة من مصادر بيانات عديدة ومتفرقة للوصول للبيانات المستهدفة" (الهادى،2019).

استعرضنا فيما سبق نماذج من التعريفات التى تمثل وجهات نظر متعددة، ويمكن أن نستخلص منها ما يلى:

- أن علم البيانات هو مجال ناشئ أو وليد بصرف النظر عن بعض الدعاوى التى ترى أنه مسمى جديد لعلم الإحصاء.
- أنه مجال بينى متعدد الارتباطات من حيث نشأته وتطوره.

- أنه يعتمد على التكنولوجيات الحديثة المتمثلة فى الطرق والمعالجات والخوارزميات والنظم.
- أنه يتعامل مع البيانات المهيكلة والبيانات غير المهيكلة.
- أن هدفه الأساسى استخراج المعرفة والبصائر أو الرؤى من البيانات لخدمة اتخاذ القرارات وشؤون أخرى.
- أنه يهتم أساسا بالبيانات الضخمة ، وأيضا البيانات البحثية.
- أن علم البيانات يغطى دورة حياة البيانات بعناصرها المختلفة، وإن كان أبرزها تحليلات البيانات.
- أن أكثر المجالات ارتباطا بعلم البيانات هى الإحصاء وعلم الحاسب وعلم المعلومات.

ولم يلمس الباحث تعريفا واضحا يمثل وجهة نظر علم المكتبات و المعلومات ، ولذلك فإنه يقدم التعريف التالى:

علم البيانات هو علم بينى متعدد الارتباطات يستخدم التكنولوجيات الحديثة من طرق ومعالجات وخوارزميات ونظم من أجل التعامل مع البيانات المهيكلة وغير المهيكلة وخاصة البيانات الضخمة بهدف استخراج أو استخلاص معرفة ورؤى لخدمة اتخاذ القرارات السليمة والمناسبة و البحث العلمى.

لكن، لماذا علم البيانات ؟ و لماذا الآن ؟ إن وراء ذلك ثلاثة أسباب :

1 . أتاحت التكنولوجيا الجديدة إمكانية التقاط و تفسير و اختزان كميات هائلة من بيانات وسائل التواصل الاجتماعى ، والدخول للانترنت ، وبعد جمع كل هذه البيانات بدأ النظر فى ماذا يمكن عمله بها.

2. أدى التقدم فى الحوسبة الى إمكان تحليل البيانات بطرق جديدة .

3. أثبتت شركات التكنولوجيا البارزة مثل جوجل و فيسبوك قوة تحليلات البيانات الحديثة (Skiena, 2017, p. 1) .

و كان تزايد الاهتمام بعلم البيانات راجعا الى :

1. الرقمنة المتزايدة و عمليات جمع البيانات التى ولّدت بيانات كثيرة .
2. الأحجام المتزايدة من البيانات و تعقدها و التى تطلبت التحليل لأسباب علمية أو

مرتبطة بالأعمال.

3. الكميات الكبيرة من قوى الحاسب المتاحة الآن سواء محليا أو عبر السحابة .
4. التزايد الواضح فى البرمجيات المتاحة .
5. السـرعة و القدرة المتزايدة للانترنت التى أتاحت الاتصـال بين أجهزة متنوعة بالاضـافة الى الأجهزة الحاسوبية المعتادة فيما يعرف ب انترنت الأشياء(Dill ; Kasik, 2019 , p. 2 ; Earnshaw) .

تبقى الاشـارة الى أن علم البيانات فى رأى البعض تنقصـه هوية مميزة و نهج نظرى أو مجموعة نظريات ملاءمة توجه دراسات علم البيانات فى حد ذاته (de Moura ,2022) ، ولذلك رأى البعض أنه من السهل تعريف علم البيانات من الناحية التطبيقية و ليس النظرية ، و ربما كان ذلك بسبب أنه مايزال فى مراحله الأولى. ومن ناحية أخرى يبدو أن المشكلة أن علم البيانات فى وضـعه الحالى ليسـت له هوية واحدة محددة متفق عليها، إذ يمكن تحديد ثلاث هويات على النحو التالى :

- **الهوية الأولى** هى ارتباط علم البيانات الوثيق بقطاع الأعمال و الصـناعة الهادف للربح ، وهى الهويـة الشـائعـة و الأكثر سـيطرة لأنها ترتبط باسـتخلاص الرؤى من البيانـات الضخمة لدى الشـركات و المؤسـسـات الصـناعية الكبيرة من أجل اتخاذ قرارات صـائبة لأصـحاب المصـلحة ، وهو ما يتطلب اسـتخدام تقنيات و منهجيات علوم مثل الاحصـاء والرياضيات و الحاسب .
- **الهوية الثانية** هى ارتباط علم البيانات بقطاع الخدمات الحكومية و ما شـابهها غير الهادف للربح و خاصـة مجال الرعاية الصحية و الرعاية الاجتماعيـة و ما الى ذلك و هو مايتطلب أيضا تحليلات البيانات لاتخاذ القرارات الجيدة .
- **الهوية الثالثة** هى ارتباط علم البيانات بقطاع البحث العلمى و خاصـة ادارة البيانات البحثية المرتبطة بالبحوث العلمية ، و هو توجه مختلف عن التوجهين السابقين .

هكذا يبدو الأمر فى الوقت الحاضـر من وجهة نظرنا ، ولعل المسـتقبل يشـهد كيانا واحدا لعلم البيانات بهوية محددة و نظم و معالجات و نظريات تُستخدم فى تطبيقات متعددة.

2. تاريخ علم البيانات

ترجع نشـأة علم البيانات إلى السـتينيات من القرن العشـرين عندما كتب جون تاكى John Tukey عـالم الرياضـيات الأمريكى مقالـة فى عـام 1962 بعنوان : The future of data analysis تناول فيها التحول فى عالم الإحصـاء ووصـف مجال أطلق عليه "تحليل البيانات" ، وهو يشبه علم البيانات الحديث.

وفى عام 1974 نشر عالم الحاسوب الدنماركى بيتر ناور Peter Naur كتابه : Concise survey of computer methods الـذى يغطى طرق معـالجـة البيـانـات عبر نطـاق عريض من التطبيقات وقدم مصـطلح data science لأول مرة معرفا له بأنه العلم الذى يتعلق بالبيانات بمجرد إنشائها بينما علاقة البيانات بما تمثله تخص مجالات وعلوم أخرى.

وفى عام 1977 أنشـئت الجمعية الدولية للتحسـيب الإحصـائى وكانت رسـالتها هى الربط بين المنهجية الإحصائية التقليدية وتكنولوجيا الحاسب الحديثة ومعرفة خبراء المجالات من أجل تحويل البيانات إلى معلومات ومعرفة.

وفى عام 1985 وفى محاضـرة ألقيت فى الأكاديمية الصـينية للعلوم فى بكين اسـتخدم جيف واى C.F. Jeff Wu مصطلح "علم البيانات" لأول مرة كبديل للاحصاء.

وفى عام 1996 تقابل أعضـاء الاتحاد الدولى لجمعيات التصـنيف فى كوب Kobe باليابان فى مؤتمرهم الذى ينعقد كل سـنتين ولأول مرة يضـمَّن مصـطلح علم البيانات فى عنوان المؤتمر: "علم البيانات، التصنيف والطرق ذات الصلة".

وفى عام 1997 دعى جيف واى C.F. Jeff Wu فى محاضـرته الافتتاحية لكرسى Carver فى الإحصـاء فى جامعة ميتشـجان بالولايات المتحدة الأمريكية إلى إعادة تسـمية الإحصـاء بعلم البيانات، وبأن يُطلق على اختصاصيي الإحصاء علماء البيانات.

وفى بـدايـة القرن الواحـد والعشـرين نشـر ويليـام كليفلانـد William Cleveland دراسـته: "علم البيانات: خطة عمل لتوسـيع المجالات الفنية للإحصـاء" ، وذكر أنه بسـبب أن الخطة طموحه وتتضمن تغييرا جوهريا فإنه يقترح تغيير المجال بأن يسمى علم البيانات.

وفى أبريل عام 2002 بدأت مجلة Data Science Journal فى الصـدور لنشـر بحوث ودراسـات عن إدارة البيانات وقواعد البيانات فى العلوم والتكنولوجيا. وتصـدر هذه المجلة عن لجنة البيانات للعلوم والتكنولوجيا المتفرعة من المجلس الدولى للعلوم. وفى يناير 2003 نشـرت

جامعة كولومبيا دوريتها Journal of data science وهى تقصد بعلم البيانات كل شئ يمكن عمله مع البيانات : الجمع، التحليل، النمذجة... ومع هذا فالجزء المهم هو التطبيقات واعتبار المجلة منصة للعاملين فى مجال البيانات لتقديم آرائهم وتبادل الأفكار.

وفى عام 2005 حددت المؤسسة الوطنية للعلوم بالولايات المتحدة علماء البيانات على أنهم : علماء المعلومات والحاسوب، مهندسو ومبرمجو قواعد البيانات والبرمجة، خبراء المجالات، المكتبيون، الأرشيفيون وغيرهم من المعنيين بالإدارة الناجحة لمجموعة البيانات الرقمية.

وفى عام 2007 نشأ مركز البحث لعلم البيانات فى جامعة فودان بالصين Fudan University . ويعقد المركز لقاءات سنوية عن علم البيانات كما نشر إثنان من الباحثين فيه هما إكسيونج Xiong وزهو Zhu كتابا بعنوان : مقدمة لعلم البيانات فى عام 2009 وذكرا أن هذا العلم يختلف عن العلم الطبيعى والعلم الاجتماعى فى أنه يأخذ البيانات فى الفضاء السيبرى كشئ للبحث، إنه علم جديد.

ومن العلامات المهمة ما كتبه دافينبورت وباتل فى سبتمبر 2012 عن عالم البيانات فى أن عمله يمثل العمل السادس فى القرن الحادى والعشرين، وقد نشرا ذلك فى دراسة ب Harvard Business Review.

وفى عام 2013 إنعقد فى لوكسمبورج أول مؤتمر أوروبى عن تحليلات البيانات وتم تأسيس الجمعية الأوروبية لعلم البيانات، وفى عام 2014 تغير اسم قسم التعلم الإحصائى وتنقيب البيانات التابع للجمعية الأمريكية للإحصاء إلى قسم التعلم الإحصائى وعلم البيانات وهو ما يعكس تصاعد علم البيانات.

ومنذ عام 2015 بزغ الذكاء الاصطناعى وتعلم الآلة والتعلم العميق و كان لها ظهور فى مجال علم البيانات . و من الأوجه المهمة فى تطور علم البيانات قدوم تنظيمات جديدة فى المجال فى عام 2018.

ومنذ عام 2020 نجد تقدما فى الذكاء الاصطناعى وتعلم الآلة فضلا عن الطلب المتزايد للمتخصصين المؤهلين فى البيانات الضخمة (شكل1) .

1962	مستقبل تحليل البيانات (ناكى Tukey)
1970	قواعد البيانات العلائقية (كودCodd)
1974	المسح الموجز لطرق الحاسب(ناورNaur)
1977	إنشاء الجمعية الدولية للتحسيب الاحصائى
1997	الاحصاء= علم البيانات (واىWu)
2001	علم البيانات : خطة عمل لتوسيع المجالات الفنية للاحصاء (كليفلاند Cleveland)
2002	إنشاء مجلة علم البيانات
2006	هادوب (ياهو)
	أول مرة يظهر علم البيانات على غلاف الاتحاد الدولى لجمعيات التصنيف
2009	ظهور عالم البيانات(يان Yun)
	البيانات المترابطة (بيرنز – لى)
2010	تاكسونومى علم البيانات(ماسون و ويجنزMason& Wiggins)
	رسمة فين لعلم البيانات (كونوايConway)
2011	بحيرات البيانات (ديكسونDixon)
	بناء فرق علم البيانات(باتلPatil)
2012	عالم البيانات :العمل السادس فى القرن 21(دافينبورت و باتل)
2018	بزوغ تكنولوجيات الرسوم
2019	بايثون واحدة من أكثر لغات البرمجة استخداما فى العالم
2020	منصات علم البيانات . الذكاء الاصطناعى

شكل رقم (1) محطات فى تاريخ علم البيانات

وهكذا فإن من العلامات البارزة اقتراح ناور فى عام 1974 لأن يحل مصطلح علم البيانات كبديل لعلم الحاسب وبالتالى فإن تقديم المصطلح يعزى إليه، ومع هذا فإنه يحسب لـ كليفيلاند أنه أنشأ علم البيانات كتخصص مستقل فى العصر الحديث (Press, 2013; Foote, 2021; History of data science, 2022; Ohri, 2021; Cao, 2017)

إن رحلة علم البيانات تكاد تكون شبيهة برحلة علم المعلومات فكلاهما بدأ فى البزوغ فى أوائل الستينيات من القرن العشرين، وكلاهما كانت نشأته بسبب عدم قدرة الطرق التقليدية التى كانت موجودة فى ذلك الوقت على التعامل أو المعالجة وتقديم الخدمات المرتبطة

بالبيانات من ناحية والمعلومات من ناحية أخرى، وكلاهما بدأ بمسميات مختلفة فقد نشأ علم المعلومات بعد تسميات كانت سائدة قبل ظهور هذا المصطلح هى التوثيق واسترجاع المعلومات والشيئ نفسه بالنسبة لعلم البيانات فقد كان المصطلح السائد فى ذلك الوقت هو تحليل البيانات، وكلاهما بينى ومتعدد الارتباطات مع اختلافات فى تحديد البينية وتعددية الارتباطات، فإذا كان علم البيانات نشأ فى أحضان الإحصاء والحاسب أساسا فإن علم المعلومات نشأ فى أحضان علم المكتبات والتوثيق ، وكلاهما استقر على أنه تخصص مستقل عن المجالات التى ظهر فيها، فقد انفصل علم البيانات عن الإحصاء و علم الحاسب وانفصل علم المعلومات عن علم المكتبات والتوثيق، وإذا كان علم البيانات قد استخدم له فى البداية datalogy ثم data science فإن علم المعلومات استخدم له فى البداية أيضا Informatology ثم Information science .

وعلى الجانب العربى ذُكر مصطلح "علم البيانات" لأول مرة فى دليل "الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات: 2016-2020، واشتملت تلك الإصدارة من الدليل على ثلاث ترجمات عربية لمقالات بالإنجليزية عن علم البيانات (عبد الهادى، 2021). ومن الكتابات العربية المهمة فى علم البيانات كتاب محمد الهادى عن ثورة البيانات وآثارها المجتمعية والتنموية والعلمية وتناول علم المعلومات فى أحد فصوله بالتفصيل (الهادى، 2019). نشير أيضا إلى مؤتمر جمعية المكتبات المتخصصة فرع الخليج العربى الذى إنعقد فى مسقط، بسلطنة عمان 6-8 مارس 2018 عن البيانات الضخمة وآفاق استثمارها، كما اهتمت بعض الدراسات العربية التى نشرت فى عامى 2019 و 2020 بتناول تدريس علم البيانات فى أقسام المكتبات والمعلومات . و نشأت " الجمعية السعودية لعلم البيانات " فى عام 2020.

وقد قدمت دينا محمد فتحى مراجعة علمية للنتاج الفكرى العربى حول البيانات وعلم البيانات، حيث تم حصر 146 دراسة أغلبها دراسات مقدمة فى مؤتمرات أو منشورة فى دوريات وكان معظمها عن البيانات الضخمة، وتبين محدودية الكتابات عن علم البيانات كمجال (فتحى، 2021).

كما تضمن عنوان مجلة عربية جديدة كلمة "بيانات " هى المجلة العربية لتكنولوجيا المعلومات والبيانات (2021).

و تجدر الاشارة الى أن المؤتمر العلمى الثانى عشر لقسم المكتبات و الوثائق و تقنية

المعلومات بجامعة القاهرة الذى انعقد فى مارس 2022 كان عن ثورة البيانات و تأثيرها على مؤسسات المعلومات العربية بين الواقع و طموحات المستقبل .

3. أهمية علم البيانات و قيمته

إن الغرض الأساسى لعلم البيانات هو إيجاد الأنماط للبيانات ، وهو يستخدم أساليب إحصائية متعددة لتحليل وسحب الرؤى من البيانات، وهو ما يساعد عالم البيانات على التفحص بعمق وعمل التنبؤات من البيانات وغرضه هو اشتقاق الاستنتاجات من البيانات ومن هذه الاستنتاجات فإنه يمكنه مساعدة الشركات وغيرها من المؤسسات فى اتخاذ القرارات المناسبة (What is the purpose of data science?)

وإذا كانت البيانات ثروة تفوق ثروة النفط كما يرى البعض فانها تحتاج إلى أن تُحلل وذلك هو اختصاص علم البيانات.

وعلم البيانات لديه إمكانية تحسين كل من مستوى الحياة ونوعية الحياة بطرق عديدة، فهو يمكن أن يساعد الناس على اتخاذ قرارات أفضل، وحل المشكلات، واكتشاف رؤى جديدة، كما أنه يمكن أن يساهم أيضا فى حل كثير من المشكلات الضاغطة على مستوى العالم.

وعلم البيانات واحد من أهم المجالات التى تتيح عملا فى القرن الواحد والعشرين وكما أشار دافينبورت و باتل إلى أن هذا العمل يحتل المرتبة السادسة فى القرن الواحد والعشرين، وهكذا فالمستقبل هو فى علم البيانات.

ويرى البعض أن تطبيق علم البيانات فى عمليات الأعمال يمكن أن يكون له مزايا عديدة للمؤسسة منها :

- المساعدة على تحسين الموارد وخفض التكاليف عن طريق إيجاد طرق أفضل لتسليم المنتجات.
- خفض الوقت ، فإن تجهيز كميات ضخمة من المعلومات فى الوقت المحدد يمكن أن يمد ب بصائر أو رؤى جديدة بشكل أسرع .
- تقديم منتجات و ابتكارات جديدة عن طريق العلم باتجاهات السوق وكشف ما يريده المستهلكون واحتياجاتهم.
- دعم القرارات، فالبيانات يمكن أن تساعد فى عمليات اتخاذ القرارات.

- قيمة الوقت الحقيقى، فإن معالجة المعلومات فى الوقت الفعلى يمكن أن يساعد فى عدم فقد الفرص فضلا عن منع مواقف أو حالات معينة غير مرغوبة.
- الدقة، فعن طريق تقنيات علم البيانات يمكن تحسين دقة الأعمال التى يتم تطويرها.
- الاختلاف عن المنافسين، فالبقاء المستمر على علم باحتياجات السوق يجعل المؤسسة فى وضع القائد، وإعطاء العملاء ما يريدون (Pisani, 2021).

وهكذا يعتبر علم البيانات ضروريا للمؤسسات المختلفة لكى تستمر فى التواجد وتتنافس فى المستقبل.

ويتغلغل علم البيانات بما يقدمه من تحليلات وبصائر وتنبؤات فى مختلف أنواع النشاط بالمجتمع حيث تتواجد تطبيقاته فى مهام التسويق المختلفة وفى إدارة علاقات العملاء وفى صناعة التمويل لتحديد درجات الائتمان وفى التجارة إلى جانب توظيفه فى عمليات الأعمال المختلفة وإدارة القوى العاملة بأى مؤسسة ، هذا فضلا عن دوره المؤثر فى ادارة البيانات البحثية فى البحث العلمى.

وإضافة إلى ما سبق فإن علم البيانات باعتباره من المجالات الجديدة الحيوية قد خلق فرصا عديدة لتقديم الاستشارات والبرامج التدريبية وورش العمل.

وعلم البيانات مطلوب لدرجة كبيرة وتوجد فرص كثيرة للعمل، وهو أسرع عمل متنامى على Linkedin ويتوقع بأنه سوف يخلق 11.5 مليون وظيفة فى 2026.

وهو يستخدم فى مجالات عديدة مثل الرعاية الصحية والبنوك والخدمات الاستشارية وصناعات التجارة الإلكترونية ومن ثم توجد فرص عمل فى مجالات عديدة ، ولعلماء البيانات وضع مهم بسبب دورهم فى تحليل البيانات واستخلاص نتائج مهمة (Pros and cons of data science) .

4. مجالات الاهتمام فى علم البيانات

ليس من السهل تحديد مجالات الاهتمام فى علم البيانات بسبب أنه مجال بينى فضلا عن أنه ما يزال فى طور النمو. وقد تبين أن مجالات التخصص فى علم البيانات المطروحة فى برنامج للماجستير على الخط المباشر فى علم البيانات فى جامعة The Pace University هى:

- تنقيب البيانات والتحليل الإحصائى : يتضمن تنقيب البيانات تحليل مجموعات كبيرة من البيانات لإنتاج معلومات ذات معنى.

- هندسة البيانات : بناء وصيانة أطر عمل تحول البيانات إلى صيغة format نافعة لأغراض التحليل ، وهذا يتضمن الجمع والتنظيف وهيكلة البيانات من مصادر مختلفة فى مستودع واحد.
- إدارة ومعمارية قواعد البيانات : انشاء حلول جديدة لكيفية تنظيم البيانات واستخدامها من قبل أصحاب المصلحة.
- هندسة تعلم الآلة : إنشاء أو تطوير نماذج نظرية يتم تغذيتها فى برمجية ذاتية التشغيل لجعل النموذج يعمل على نطاق واسع.
- ذكاء واستراتيجية الأعمال : تحليل البيانات وتطوير بصائر يمكن أن تساعد فى تحسين أداء الأعمال.
- التمثيل المرئى للبيانات : عرض البيانات باستخدام الأدوات المرئية التفاعلية مثل : الرسومات، الأشكال، الانفوجرافيك.
- تحليل بيانات العمليات : تحديد مجالات التحسين فى عمليات الأعمال باستخدام البيانات المقدمة من الأعضاء الآخرين فى فريق علم البيانات ثم استخدام برامج إحصائية لتقييم الحلول العملية لمشكلات الأعمال وتقديم النصيحة للمديرين لأفضل ما يمكن عمله.
- تحليل بيانات التسويق : ممارسة دراسة البيانات لقياس وتحسين فاعلية الحملات التسويقية (Eight data science specializations).

ويركز المنهج البؤرى لدرجة الماجستير على الخط المباشر فى علم البيانات فى المدرسة الذكية لعلم المعلومات بجامعة كاليفورنيا ، بيركلى بالولايات المتحدة على المهارات المفتاحية التالية:

- تصميم البحث.
- هندسة البيانات.
- التمثيل المرئى للبيانات.
- التحليل الإحصائى.
- تنظيف البيانات.
- تنقيب البيانات واستكشافها.

- أخلاقيات المعلومات والخصوصية.
- تعلم الآلة (Online Master's in data science) .

وهناك من يذكر أن مجالات علم البيانات التخصصية هى فى الوقت الحاضر على النحو التالى :

- الإحصاء والاحتمالات.
- باثيون.
- تعلم الآلة.
- معالجة البيانات.
- التمثيل المرئى للبيانات.
- تنقيب البيانات.
- التحليلات التنبؤية.
- البيانات الضخمة.
- النمذجة.
- استشارات البيانات (10 top data science fields).

وقد قام كل من وانج Wang و جو Joo وليو Lu باستكشاف الموضوعات فى مجال علم البيانات من وثائق الويكيبديا اعتمادا على تجميع تحليل المكونات الأساسية و نمذجة الرؤوس و انتهوا الى أن هناك ست مجموعات تشكل موضوعات مجال علم البيانات على النحو التالى :

المجموعة الأولى : تنقيب البيانات و البيانات الضخمة ، وهى تضم : تنقيب البيانات ، التمثيل المرئى للبيانات ، البيانات الضخمة ، تحليل البيانات ، ذكاء الأعمال .

المجموعة الثانية : علم المعلومات و علم الحاسب .

المجموعة الثالثة : معالجة اللغة الطبيعية و اكتشاف المعرفة .

المجموعة الرابعة : تعلم الآلة.

المجموعة الخامسة : الطرق و التقنيات و التطبيقات ، مثل : الحفظ الرقمى ، الميتاداتا، حزم البرمجيات مفتوحة المصدر، التحليل التنبؤى .

المجموعة السادسة: علم الويب ، و يتضمن : الشبكة العنكبوتية العالمية ، زاحف الويب ،

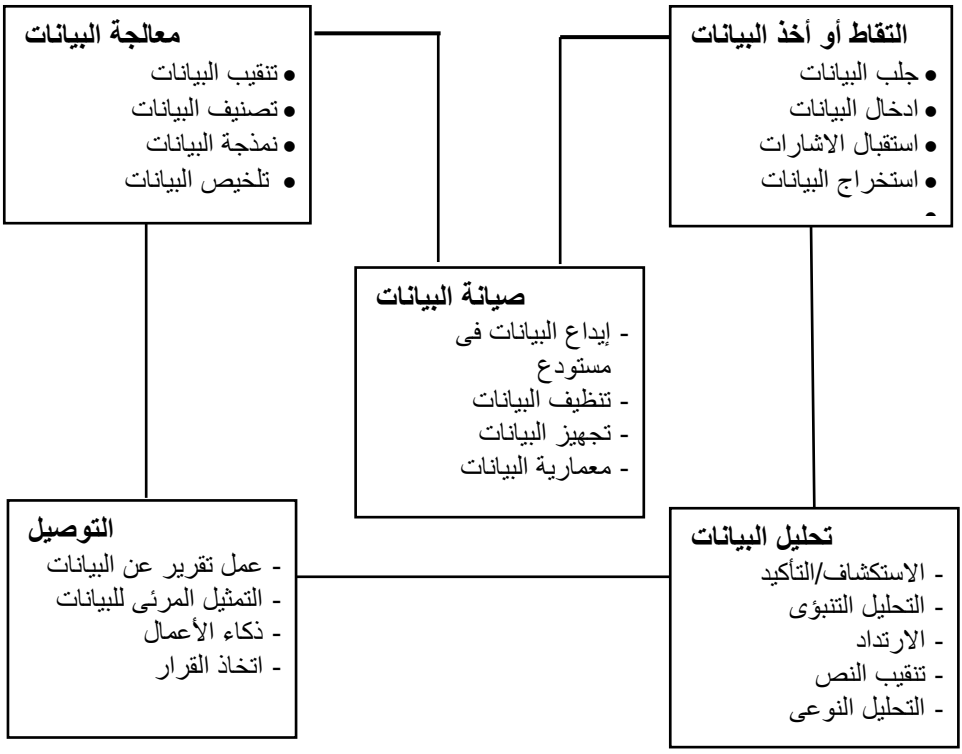
قائمة مبادرات أرشفة الويب ، الويب الاجتماعى، مواقع الويب (Wang;Joo; Lu, 2014) .

وبين سونج وزهو أن الركائز الأساسية لعلم البيانات هى:

- البيانات: المهيكلة وغير المهيكلة.
- التقنيات: مثل أنظمة هادوب، تقنيات تنقيب البيانات، تقنيات التعلم الآلى.
- الأفراد: علماء الحاسب والإحصاء وخبراء المجال وعلماء البيانات ومحللى الأعمال (Song; Zhu, 2016) .

وتشير دراسات عديدة إلى دورة حياة علم البيانات على الوجه التالى:

- التقاط أو أخذ البيانات وتضم: جلب البيانات، إدخال البيانات، استقبال الإشارات، استخراج أو استخلاص البيانات . وتتضمن هذه المرحلة جمع البيانات المهيكلة وغير المهيكلة.
- صيانة البيانات وتضم: ايداع البيانات فى مستودع ، تنظيف البيانات، تجهيز البيانات، معمارية البيانات. وتتضمن هذه المرحلة وضع البيانات الخام فى شكل يمكن استخدامه.
- معالجة البيانات وتضم: تنقيب البيانات، التصنيف، نمذجة البيانات، تلخيص البيانات . وتتضمن هذه المرحلة أخذ البيانات المجهزة وفحص أنماطها ومداها وتحيزاتها لتقرير مدى نفعها أو فائدتها فى التحليل التنبؤى.
- التحليل وتضم الاستكشاف /التأكيد أو التثبيت،التحليل التنبؤى،الارتداد، تنقيب النص، التحليل النوعى. و تتضمن هذه الحلقة إنجاز مختلف التحليلات على البيانات، وهى أهم الحلقات.
- التوصيل ويضم تقديم تقارير عن البيانات، العرض المرئى للبيانات، ذكاء الأعمال، اتخاذ القرار. وتتضمن هذه الحلقة الأخيرة تجهيز أو إعداد التحليلات فى صيغ قابلة للقراءة بسهولة مثل الرسومات والتقارير (What is data science : life cycle) (شكل 2).



شكل رقم (2) دورة حياة علم البيانات

(المصدر: ischoolonline. berkeley.edu/data-science)

و قد تم اختيار كتابين من الكتب عن علم البيانات الموجهة لاختصاصيى المكتبات و المعلومات لعرض محتوياتهما، و قد جاءت فصول الكتاب الأول : Data science for librarians على النحو التالى :

1. بيانات أكثر، مشكلات أكثر.
2. ساحل جديد للمكتبيات .
3. إنشاء البيانات وجمعها.
4. البيانات لاختصاصي المكتبة الأكاديمية.

5. خدمات البيانات البحثية .
6. مصادر البيانات .
7. العناية بالبيانات (الأرشفة/ الحفظ) .
8. اختزان البيانات وادارتها واسترجاعها .
9. تحليل البيانات والتمثيل المرئى لها.
10. أخلاقيات البيانات وسياساتها.
11. البيانات للمكتبات العامة والمتخصصة.
12. خاتمة: المكتبة، المعلومات وعلم البيانات (Du;Khan,2020).

ويشـــتمل الكتاب الثانى Practical data science for information professionals على الفصول التالية:

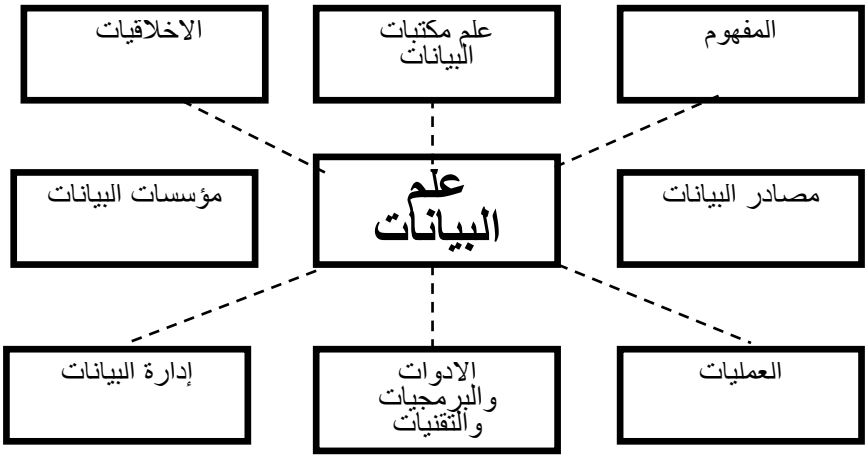
1. ما هو علم البيانات؟.
2. بيانات قليلة، بيانات ضخمة.
3. عملية علم البيانات.
4. أدوات تحليل البيانات.
5. التجميع وتحليل الشبكات الاجتماعية.
6. التنبؤات والتوقعات .
7. تحليل النص وتنقيبه.
8. مستقبل علم البيانات واختصاصيي المعلومات.
9. ملحق: مفاهيم البرمجة لعلم البيانات (Stuart,2020).

اســتعرضــنا فيما ســبق مكونات بعض برامج الماجســتير فى علم البيانات، وتوجهات بعض الباحثين بخصـــوص مجالات علم البيانات، ودورة حياة علم البيانات، ومحتويات أبرز الكتب الدراســـية الموجهة لاختصـــاصـــيي المكتبات و المعلومات ، ومن ثم فإنه يمكن اقتراح المكونات الرئيسـية لعلم البيانات تلك المناسـبة لاختصـاصـيي المكتبات و المعلومات على الوجه التالى:

- مفهوم البيانات وعلم البيانات
- عمليات علم البيانات وتشمل: إنشاء البيانات وجمعها، اختزان البيانات وأرشفتها، تحليل

البيانات، التمثيل المرئى للبيانات.

- أدوات وبرمجيات وتقنيات علم البيانات.
- إدارة البيانات الضخمة والبحثية
- مؤسسات البيانات واختصاصيي البيانات
- أخلاقيات البيانات
- علم مكتبات البيانات (شكل 3) .



شكل(3)مكونات علم البيانات

5. علاقة علم البيانات بالمجالات الأخرى

عادة ما يوصف علم البيانات بأنه علم بينى متعدد الارتباطات، وهى سمة من سمات العلوم الجديدة الناشئة، ومن ثم فعلم البيانات كتخصص مستقل له علاقة نشوء وعلاقة ارتباط بمجالات أخرى ذات صلة وثيقة به وخاصة علم الإحصاء، وعلم الحاسب، وعلم المعلومات ، والمجالات المعرفية المتخصصة.

أول المجالات هو علم الإحصاء حيث أن النظرية الإحصائية وأساليبها تسهم فى علم البيانات من خلال اتخاذ قرارات حول البيانات التى يجب جمعها أو تضمينها فى التحليل وتقديم أساليب إحصائية وتقييم فاعلية النتائج.

ويسهم علم الحاسب من خلال إنشاء أو تطويع خوارزميات رياضية وإحصائية تيسر معالجة البيانات عالية السرعة وكبيرة الحجم، ويعمل علماء الحاسبات على نمذجة توزيعات

البيانات وأنماطها باستخدام أساليب مثل تعلم الآلة كما يسهم علم الحاسب فى علم البيانات من خلال توفير أساليب للآلات والخوارزميات التى يمكن لخبراء المجالات تطبيقها فى طرح الأسئلة والإجابة عنها.

ونأتى إلى علم المعلومات حيث تبين أنه يوجد نوع من العلاقة بين كل من التخصصين فهما يشكلان معا عناصر بحث ما يُعرف بسلسلة البيانات، المعلومات، المعرفة، ويتداخل كلا التخصصين فى رسالتيهما إلى حد كبير، و يمتلك علم المعلومات المزيد من العناصر التى يمكن مشاركتها مع علم البيانات، كما أن مفهوم البيانات متأصل من حيث علاقته المتداخلة مع مفهوم المعلومات، ومن ناحية أخرى فإن علماء المعلومات يهتمون بالدورة الكاملة لحياة البيانات.

وهناك علاقة بين علم البيانات والتخصصات المعرفية الأخرى لأن مجالات المعرفة الإنسانية بوجه عام تمثل بيئات متفاوتة لتطبيقات علم البيانات فكل المجالات عادة ما تنتج بيانات مختلفة تحتاج إلى تحليل ومعالجة، ومن تلك التخصصات يبرز إدارة الأعمال والعلوم الصحية والتجارة الإلكترونية والخدمات الحكومية وغير ذلك من المجالات طالما أن علم البيانات يتعامل مع البيانات التى تنتجها القطاعات المختلفة بكميات ضخمة وتتطلب تحليلا ومعالجة بالطرق الإحصائية والرياضية التى يقدمها علم البيانات (شكل4) (وانج، 2019، ص 317-319؛ زكريا ، 2021، ص 24-27).

شكل (4) علاقة علم البيانات بالعلوم الأخرى

خلاصة

توجد تعريفات عديدة لعلم البيانات من منظورات مختلفة، و مع هذا يمكن القول أن علم البيانات هو علم بينى متعدد الارتباطات يستخدم التكنولوجيات الحديثة من طرق ومعالجات وخوارزميات ونظم من أجل التعامل مع البيانات بأنواعها المختلفة وخاصة البيانات الضخمة، من أجل استخراج أو استخلاص معرفة وبصائر أو رؤى لخدمة اتخاذ القرارات المناسبة.

وترجع نشأة علم البيانات إلى عام 1962 فيما كتبه جون تاكى عن مستقبل تحليل البيانات كما قدم تاور مصطلح علم البيانات فى عام 1974، ورأى جيف واى فى عام 1985 استخدام مصطلح علم البيانات كبديل للإحصاء، كما أن كليفلاند فى أوائل القرن 21 نشر دراسة مهمة عن خطة عمل لعلم البيانات ، وتوالت بعد ذلك إنجازات مهمة مثل إنشاء مراكز بحوث فى المجال وبدء مجلات متخصصة وعقد مؤتمرات مهمة، وتمثل الأمر مؤخرا فى الانتفاع من الذكاء الاصطناعى وتعلم الآلة وما تؤديه من دور مهم فى تطوير العلم. ويلاحظ أن الاسهامات العربية محدودة فى هذا المجال، كما يلاحظ أن رحلة علم البيانات تكاد تكون مشابهة لرحلة علم المعلومات. ولعلم البيانات أهميته وقيمته فى خدمة الأفراد والمؤسسات والمجتمع، كما تبين أن علماء البيانات يمثلون المرتبة السادسة من بين وظائف القرن 21، أما مجالات الاهتمام فى علم البيانات فقد تبين أنها ترتبط بصفة عامة بدورة حياة علم البيانات المتمثلة فى التقاط البيانات وصيانتها وتجهيزها وتحليلها وأخيرا توصيلها.

وعموما فإن علم البيانات له علاقة نشوء وارتباط بعدة مجالات أهمها الإحصاء وعلم الحاسب وعلم المعلومات، كما تبين انتفاع مجالات معرفية عديدة من طرق وأساليب علم البيانات.

المصادر

زكريا، محمود شريف (2021). تعليم علم البيانات فى المنطقة العربية.- المجلة العلمية للمكتبات والوثائق والمعلومات.- مج2، ع 6 (أبريل).- ص 24-27.

عبد الهادى، محمد فتحى (2021). الإنتاج الفكرى العربى فى مجال المكتبات والمعلومات: 2016-2020م.- تونس: الاتحاد العربى للمكتبات والمعلومات.

فتحى، دينا محمد (2021). البيانات وعلم البيانات فى ضوء النتاج الفكرى العربى: دراسة تحليلية.- الفهرست.- ع71 (يناير).- ص 203-250.

معجم البيانات والذكاء الاصطناعى: إنجليزى- عربى (2022).- الرياض: الهيئة السعودية للبيانات والذكاء الاصطناعى.- 148ص.

الهادى، محمد محمد (2019). ثورة البيانات وآثارها المجتمعية والتنموية والعلمية.- القاهرة: روابط للنشر وتقنية المعلومات.

وانج، لين (2019). توأمة علم البيانات مع علم المعلومات بمدارس علم المكتبات والمعلومات/ ترجمة محمد إبراهيم حسن الصبحى.- المجلة العلمية للمكتبات والوثائق والمعلومات.- مج1، ع2 (يوليو).- ص311-339.

10 top data science fields. www.datasciencedegreeprograms.net

Cao, Longbing (2016). Data science: a comprehensive overview.- ACM Computing Surveys.- Vol.50, No.3, p.1-42.

Cervon, H.Frank (2016). Informatics and data science: an overview for the information professional. - Digital Library Perspectives.- Vol.33, No.1.- p.7-10.

Data Science - Cambridge English Dictionary. https://dictionary.cambridge.org

Data science- Wikipedia. https://en.wikipedia.org

Data science definition. https://techterms.com

de Moura , Pedro Jacome(2022). Is data science a science ? . — Kybernetes. — Vol. 51, No. 7. — p. 2416-2434.

Du, Yunfei; Khan, Hammad (2020). Data Science for Librarians.- Santa Barbara, Calif.: Libraries Unlimited.

Earnshaw, Rae ; Dill, John; Kasik, David(2019).Data science and visual computing. Springer.

Eight data science specializations (2021). https://www.kdnuggets.com

Foote, Keith (2021). A brief history of data science. www.dataversity.net

Herndon, Joel (2022). Introduction: The rise of data science. In: Herndon, Joel (ed.). Data science in the library. - London: facet publishing.

History of data science (2022). https://www.careerera.com

NIST(2015). Big data interoperability framework. Vol.1. http://bigdatawg.nist.gov

Ohri, Ajay (2021). History of data science. www.jigsawacademy.com

Online Master's in data science. http://ischoolonline.berkeley.edu/data-science

Patil, D. (2015). A memo to the American people from US chief data scientist dr. D.J. Patil. https://obamawhitehouse.archives.gov/blog

Pisani, Mikaela(2021). The value and advantages of data science . https://www.rootstrap.com

Press, Gil (2013). A very short history of data science. www.forbes.com

Pros and cons of data science. http://data-flair.training

Provost, F.; Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. - Big Data. — Vol.1,No.1. -.p. 51-59.

Skiena, Steven (2017). The data science design manual. Springer.

Song, Il-Yeol; Zhu, Yongjun (2016). Big data and data science. - Expert Systems, Vol.33, No.4, p.364-373.

Stuart, David (2020). Practical data science for information professionals. - London: facet publishing.

Wang, Yanyan;Joo, Soohyung;Lu,Kun(2014).Exploring topics in the field of data

science by analyzingWikipedia documents.

77 th ASIS&T Annual meeting,2014.- Seattle, WA, US. - 4p.

What is data science? -Amazon AWS. https://aws.amazon.com

What is data science? - IBM. https://www. ibm.com

What is data science? (2021). https://www.techopedia.com

What is data science & how does it work? (2022). https://mygreatlearning.com

What is data science: Lifecycle, applications, prerequisites and tools (2022). https://www.simplilearn.com

What is the purpose of data Science?. https://data-flair.training

Data science concept

Dr. Mohamed Fathy Abdelhady
Professor of Library and Information science
Faculty of Arts – Cairo University

Abstract

Data science is one of the modern sciences, and it has grown and flourished very quickly in the past few years due to its association with many scientific fields on the one hand and its benefit from the successive developments in the field of information and communication technology on the other hand. This study seeks to introduce data science from multiple perspectives in order to derive a definition that represents the viewpoint of library and information science. The study dealt with data science history and development over about fifty years, as well as its value as a vital field at the present time, and its topics covering the life cycle of data, and its relationship with other fields.

keywords: Data science; Data; Information science; Data science areas of interest; Data life cycle.