علم البيانات Data Science
علم المستقبل

أ.د. ربحي مصطفى عليان

أستاذ علم المكتبات والمعلومات

كلية العلوم التربوية- الجامعة الأردنية

المستخلص:

يعد علم البيانات من العلوم الجديدة التي ظهرت مع بدايات القرن الحادي والعشرين في الدول المتقدمة وبخاصة الولايات المتحدة الامريكية التي تقوم بتدريسه على مستوى البكالوريوس والماجستير والدكتوراة، تقدم الدراسة مقدمة عامة لهذا التخصص الجديد ومفهومه، وعالم البيانات والمهارات التي يجب أن يتقنها، والاختلافات بين محلل البيانات وعالم البيانات، وايجابيات وسلبيات دراسة تخصص علم البيانات، وأشهر الجامعات التي تقوم بتدريسه، وبرنامج مايكروسوفت المهني في علم البيانات.

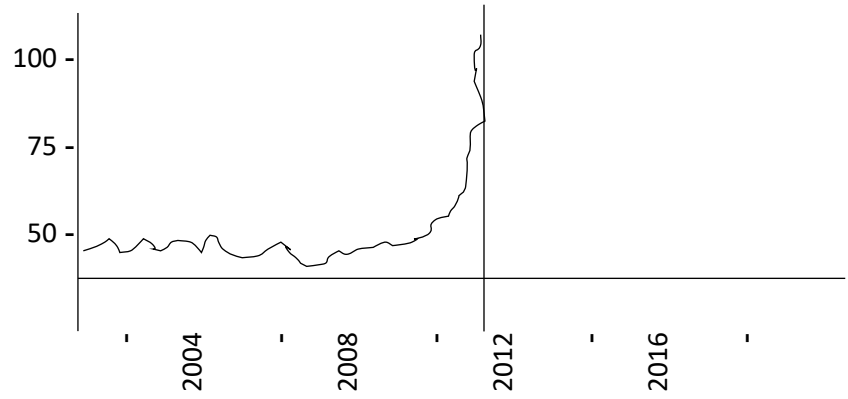
الكلمات المفتاحية: علم البيانات؛ علماء البيانات؛ محللو البيانات؛ برامج علم البيانات.

تاريخ استلام البحث: 2022/5/10 **تاريخ إجازة البحث: 2022/6/12**

مقدمة عامة:

مصطلح علم البيانات (Data Science) والذي استخدم في الأصل بالتبادل مع (Data logy) موجود لأكثر من 30 عاماً، وكان يستخدم في البداية كبديل عن علم الحاسوب (Computer Science). وقد استخدم بيتر نادر عام 1960م، الذي نشر في عام 1974 مسحاً موجزاً لأساليب الحاسوب، واستخدم علم البيانات في دراسة لطرق معالجة البيانات المعاصرة التي تستخدم في مجموعة واسعة من التطبيقات (Wikipedia, 2020).

وعندما أطلقت مجلة (Harvard Business Review) على علم البيانات عام 2012م، على أن الوظيفة الأكثر إثارة في القرن الحادي والعشرين، أصبح مصطلح علم البيانات عبارة شائعة. ويستخدم الآن بصفة تبادلية مع تحليل الأعمال وذكاء الأعمال، والتحليل التنبؤي، والإحصاء. ويرى هانز روزلينغ أن علم البيانات قد جعل من الإحصاء أكثر إثارة. وقد أعيد تسمية العديد من المناهج السابقة لعلم البيانات لكي تصبح أكثر جاذبية. غير أنه لا يوجد توافق حول المناهج المناسبة، وقد تزايد الاهتمام بعلم البيانات في نهاية العام 2012م بشكل ملحوظ سواء من قبل الشركات أو من قِبل الأفراد الراغبين بتعلم هذا العلم الجديد. ويُظهر الشكل (1) المأخوذ من Google Trends مدى ازدياد البحث عن مصطلح علم البيانات ما بعد عام 2012م، ومع تعدد المصادر التي تتولد منها البيانات وإدراك الشركات لمدى أهميتها، يتوقع تزايد الاهتمام بهذا العلم في السنوات القادمة.



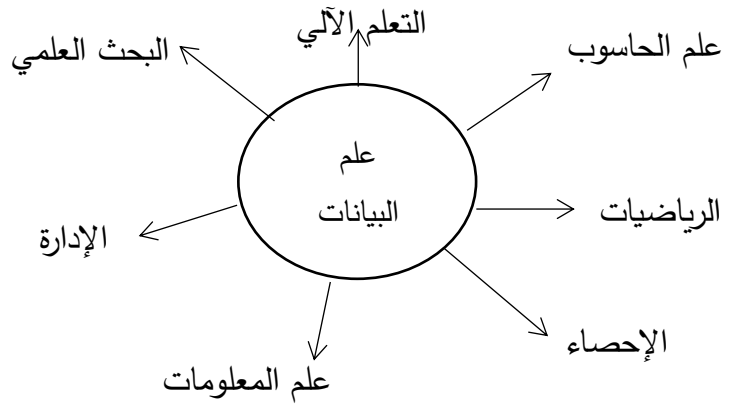
شكل (1) ازدياد البحث عن علم البيانات

ويرى جيم غراي الفائز بجائزة تورينغ، أن علم البيانات نموذج رابع للعلوم (التجريبية والنظرية والحسابية). وأكد أن كل شيء عن العلوم يتغير بسبب تأثير تكنولوجيا المعلومات وثورة البيانات (Tolle, 2009).

وقد أصبحت التخصصات ذات العلاقة بالبيانات تتصدر قائمة التخصصات التي يزداد الإقبال عليها من قبل الطلبة، ومن هذه التخصصات: هندسة البرمجيات ونظم المعلومات الحاسوبية وعلم الروبوتات، والذكاء الاصطناعي، وعلم البيانات الذي يعد من تخصصات المستقبل الواعدة والمطلوبة.

ويمكن حصر المقررات الأساسية لتخصص علم البيانات فيما يلي:

- الرياضيات وبشكل خاص الإحصاء والجبر الخطي.
- لغات البرمجة ومنها Matlab وPython وغيرها.
- تحليل البيانات ويعد من أهم المحاور الرئيسة للتخصص.
- إدارة البيانات وتشمل استخراج البيانات والبحث عنها واسترجاعها.
- البيانات الضخمة ويحتاج التعامل معها إلى أساليب وتقنيات حديثة.
- التعلم الآلي الذي يختص بالتعامل مع العديد من الموضوعات، مثل الروبوتات وغيرها.
- الخوارزميات التي تتكون من مجموعة من الخطوات المنطقية والرياضية للاستعانة بها في حل المشكلات.



شكل (2) الموضوعات ذات العلاقة بعلم البيانات

مفهوم علم البيانات:

علم متعدد المجالات، وهو العلم الذي يستخدم الأساليب العلمية والعمليات والخوارزميات والأنظمة بغرض استخراج البيانات والأفكار من البيانات، سواء كانت هذه البيانات منظمة أم غير منظمة. وكان يطلق عليه قديماً (Datalogy) ويعرف أيضاً بأنه: استخدام أجهزة الحاسوب ومعداته وأنظمته وبرمجياته والخوارزميات بهدف حل المشكلات، وتفسير الظواهر، وذلك بالاستفادة من نظريات مستمدة من علوم أخرى مثل الرياضيات ونظم المعلومات والإحصاء وعلم الحاسوب (علم البيانات، 2020).

ويرى البعض أن علماء البيانات يعتبرون كائنات خرافية نادرة، تتمتع بطيف واسع جداً جداً من المهارات، ويتعامل هؤلاء الخرافيون مع البيانات الأولية ليجعلوا منها مواداً يعمل عليها محللو البيانات، ويمكن أن يقوموا بذلك باستخدام تقنيات مختلفة مثل (باثيون وR) وغيرها. وعليه يتعلق عملهم بتنظيف البيانات، وهذه مهمة أساسية في العمل مع البيانات الضخمة. ومن المهارات التي يتميز بها عالم البيانات: التعلم الآلي والإحصاء والنمذجة وتقنيات عرض البيانات.

و علم البيانات مجال واسع وخليط من عدة مجالات مرتبطة مع بعضها البعض، يركز بشكل أساسي على معرفة وفهم البيانات التي تمتلكها الشركات والمؤسسات واستخدامها لحل المشكلات أو الإجابة عن أسئلة معينة أو تقديم توصيات ونصائح للإدارة لتحسين العمل أو تجنب المشكلات، وذلك باتباع المنهج العلمي (الخياط، 2018).

وهو مفهوم قائم على توحيد الإحصاء، وتحليل البيانات، وتعلم الآلة، والوسائط المرتبطة بها، من أجل فهم وتحليل ظاهرة فعلية باستخدام البيانات، ويقوم بتوظيف نظريات وأساليب مستمدة من حقول معرفية متعددة مثل :الرياضيات والإحصاء وعلم المعلومات وعلم الحاسوب. (ويكبيديا، 2020).

و هذا العلم يغلب عليه تداخل الاختصاصات، ويقوم على استخدام الأساليب العلمية، والمعالجات والخوارزميات، والنظم، لاستخراج المعرفة والأفكار من البيانات بشكليها (المهيكلة وغير المهيكلة)، بشكل مشابه للتنقيب في البيانات، ويعتمد على تقنيات تعلم الآلة والذكاء الصناعي وبرامج معالجة البيانات الضخمة (ويكبيديا، 2020).

المهام الأساسية لعالم البيانات Data Scientist :

1. **التنبؤ بقيم معينة بناءً على بيانات قديمة، مثال:**
   - التنبؤ بإجمالي الربح الذي تحصل عليه الشركة في الشهر القادم.
   - التنبؤ بعدد المستخدمين الجدد لتطبيق معين في شهر معين.
2. **عمليات التصنيف Classification، مثل:**
   - تصنيف البريد الوارد بأنه هام أو غير هام.
3. **بناء نظم التوصية Recommendation Systems، مثال:**
   - عندما تقوم بعملية شراء لمنتج ما، سوف تلاحظ وجود قائمة تخبرك بمنتجات شبيهة بالمنتج الذي اخترته.
   - أنظمة التوصية في مواقع مشاهدة الأفلام السينمائية.
4. **الكشف عن الأنماط: Pattern Detection، مثال:**
   - اكتشاف العلاقات بين البيانات الموجودة التي لا يمكن الكشف عنها بالطرق التقليدية.
5. **التحليل الاستكشافي للبيانات: Exploratory Data Analysis، مثال:**
   - إجراء بعض العمليات على البيانات لفهم طبيعتها والحصول على فكرة أولية عما تحتويه من معلومات.
6. **قياس جودة أداء الخوارزميات وتحسين نتائجها. (Zhn, 2019).**

ويقوم علم البيانات بشكل عام على ما يلي:

1- الحصول على البيانات المتعلقة بالموضوع أو المشكلة المراد حلّها.
2- معالجة البيانات والتأكد من صحتها وخلوها من الأخطاء التي من الممكن أن تؤثر في النتائج.
3- القيام بالتحليل الاستكشافي للبيانات للحصول على نظرة شاملة للبيانات المتعامل معها.
4- بناء وتطوير وتطبيق الخوارزميات والنماذج اللازمة لحل المشكلة، والوصول إلى النتائج.

5- عرض النتائج النهائية على متخذي القرار. (تعرف على علم البيانات، 2018).

أما خطوات تطبيق علم البيانات في مجال الأعمال فهي كما يلي:

- **أولاً:** طرح الأسئلة وتحديدها (المشكلة).
- **ثانياً:** جمع البيانات من مصادرها المختلفة (الداخلية والخارجية).
- **ثالثاً:** معالجة وتنظيم البيانات بالطرق المناسبة.
- **رابعاً:** تحليل البيانات بالطرق المناسبة.
- **خامساً:** تطوير النماذج والخوارزميات المناسبة.
- **سادساً:** عرض البيانات من خلال الرسوم البيانية وغيرها.
- **سابعاً:** اتخاذ القرار المناسب.

ما هو عالم البيانات:

إن ظهور البيانات الضخمة أى البيانات الكبيرة جداً والمتباينة والمتكاثرة والتي لا تستطيع تطبيقات البيانات التقليدية التعامل معها، أدى إلى زيادة الحاجة إلى المهنيين المهرة القادرين على التنقيب عن البيانات وتنظيمها وتفسيرها وتقديمها للشركات لمساعدتها في اتخاذ القرارات الاستراتيجية. وهناك 56 ألف وظيفة سنوياً في بريطانياً لوحدها للمتخصصين في علم البيانات. ولكن من هو عالم البيانات؟

عالم البيانات يقوم بأخذ البيانات الخام من مصادرها المختلفة، ثم يقوم بتحليلها ليجعلها قابلة للاستخدام وأكثر قيمة للمؤسسات. وللقيام بذلك فهو بحاجة إلى مزيج من المهارات، وإلى أسس قوية في الرياضيات والخوارزميات، وفهم جيد للسلوك البشري والقطاعات التي يعمل فيها، لوضع النتائج في السياق المناسب، واستكشاف رؤى من هذه البيانات لتحديد الاتجاهات. (الجعيد، 2020).

أما المهارات التي يجب أن يتقنها عالم البيانات فهي:

أولاً: المهارات التحليلية:

وتعني القدرة على الحصول على البيانات الخام وتحديد الأهداف، ثم وضع التحليل في السياق المناسب لإيجاد الحلول والوصول إلى توصيات يمكن تقديمها للجهات ذات العلاقة.

ثانياً: مهارات العرض:

وتعني القدرة على عرض النتائج على أصحاب العلاقة بوضوح وفعالية والقدرة على الاستماع إلى متطلباتهم وفهمها وفهم مشكلاتهم من أجل حلها.

السمات الشخصية للمتخصصين في علم البيانات:

يجب على الذين يرغبون في الالتحاق ببرامج علم البيانات التمتع بالصفات والسمات الشخصية التالية:

- الإبداع والفضول وحب الاستطلاع.
- الدقة والاهتمام بالتفاصيل.
- مهارات التواصل والتحليل والتخطيط والتنظيم وحل المشكلات.
- مستوى جيد من الرياضيات والتقنيات الكلية.
- الرغبة في تطوير النظم والبرامج والتطبيقات ومواكبة التكنولوجيا وتحدياتها.
- مهارات الفهم والتفكير والتحليل الناقد والمنطقي.
- مهارات جمع البيانات والمعلومات بالطرق المختلفة وربطها مع بعضها البعض.
- القابلية لتخزين الكثير من المعلومات وتعلم المعلومات الجديدة.
- مهارات صنع القرار واتخاذه.
- القدرة على تعلم اللغة الإنجليزية وإجادتها.

هذا بالإضافة إلى أخلاقيات العمل.

وينقسم علماء البيانات إلى قسمين رئيسين:

1- علماء بيانات يقومون ببناء نماذج رياضية (إحصائية) للبيانات الأولية وذلك من أجل الحصول على معلومات منها.
2- علماء بيانات يقومون ببناء نماذج باستخدام مهارات وبيئات برمجية (بابثون وR وغيرها، وهؤلاء يطلق عليهم اسم (محللي برمجة التطبيقات).

ويمكن للمتخصص في علم البيانات العمل في مجال البرمجة وتحليل النظم وتصميمها وتطويرها، وتحليل البيانات، والإشراف على أعمال الحاسوب والروبوتات، وفي مجال

قواعد البيانات وفي مجال هندسة الحاسوب وإدارة وحماية الشبكات، بالإضافة إلى مجال التدريب والتعليم.

وتتعدد مجالات العمل للمتخصصين في علم البيانات لتشمل القطاعات الحكومية والقطاعات الخاصة بشكل عام، والوزارات والجامعات والمكتبات ومراكز البحوث والدراسات الشركات والبنوك والمؤسسات الصناعية والمهنية والمدارس بشكل خاص، بالإضافة إلى مجال التدريس والتدريب في الكليات والجامعات ومعاهد التدريب المختلفة.

وهناك الكثير من المهارات التي يجب أن يتقنها عالم البيانات وأهمها:

- الرياضيات.
- لغات البرمجة وأهمها (Python و R وSRS).
- التعلم الآلي Machine Learning.
- التعامل مع قواعد البيانات.

ويرى (غلاس دور) أنه وحتى نهاية 2019م، فإن هناك 10 لغات وتقنيات برمجة أساسية يجب أن يتقنها عالم البيانات سواء كان يعمل في مجال البيانات الأساسية أو البحث أو البيانات الضخمة، وهي:

- لغة باثيون Python.
- لغة آر R.
- لغة إس كيوال SQL.
- لغة هادوب.
- لغة جافا Java.
- لغة ساس SAS.
- لغة سبارك.
- لغة ماتلاب Matlab.
- لغة هايف.
- لغة بولورويز.

الاختلافات الرئيسة بين محلل البيانات وعالم البيانات:

أولاً: محلل البيانات يركز بشكل أساسي على معالجة وتحليل البيانات واستخراج معلومات منها تفيد الأعمال في عملية تفسير الماضي والحاضر. أما عالم البيانات فيقوم بتطوير وبناء خوارزميات ونماذج لتحليل البيانات واستخراج معلومات منها تفيد في تفسير الماضي والحاضر (والتنبؤ بالمستقبل).

ثانياً: محلل البيانات يقوم بالإجابة عن أسئلة محددة طرحت من قبل أشخاص آخرين في معظم الأحيان. أما عالم البيانات فيقوم باكتشاف وصياغة الأسئلة متبعاً المنهج العلمي، والتي تهدف إلى حل مشكلات غير ظاهرة عن بعد.

ثالثاً: محلل البيانات عادة ما يقوم باستخدام برامج وتطبيقات جاهزة لتساعده في جميع أعماله.

أما عالم البيانات، فبالإضافة إلى استخدامه للأدوات والبرمجيات، فهو قادر على بناء أنظمة وتطبيقات خاصة به. (Zhu, 2015).

(محلل البيانات) يهتم بفهم وشرح الماضي	(عالم البيانات) يهتم بتحسين المستقبل
• - إعداد تقرير شهري بأكثر الأفلام والمسلسلات مشاهدة على الموقع. • - جمع وتحليل البيانات الخاصة من استبانات رضا المستفيدين. • - تحليل المناطق التي يتواجد فيها عدد كبير من المستخدمين وإبلاغ قسم التسويق بنتائج التحليل. • - إعداد التقارير الخاصة بمؤشرات لأداء الرئيسة الخاصة بالشركة.	• التنبؤ بعدد مشاهدات فيلم معين قبل طرحه. • بناء نظام توصية للأفلام. • اكتشاف أنماط وعلاقات في البيانات الموجودة باستخدام خوارزميات واستخدام هذه النماذج في تحسين خدمات الموقع. • تطبيق خوارزميات التصنيف لتحليل الصور الموجودة في الفيلم وتصنيفه في إحدى الفئات (يحتوي مشاهد سيئة، لا يحتوي مشاهد سيئة) وفي حال كان عمر المشاهد غير مناسب، يتم تنبيهه.

إيجابيات وسلبيات دراسة تخصص علم البيانات:

أولاً: الإيجابيات.

- تصبح متخصصاً في علم البيانات.
- القدرة على استخدام مجموعة متنوعة من التقنيات.
- القدرة على تحليل البيانات.
- القدرة على إجراء الإحصاءات المختلفة.
- القدرة على حل المشكلات باستخدام البيانات.
- يعد التخصص من أكثر التخصصات تطوراً وانتشاراً في هذه المرحلة.
- القدرة على التفكير المنطقي والعمل بروح الفريق.
- تطوير الكفاءات الشخصية في مجال التنقيب عن البيانات وخزنها واسترجاعها.
- القدرة على إعداد البحوث بالمنهج العلمي.
- فتح الباب أمام المسارات المهنية والأكاديمية في المجال.
- الاستفادة من تكنولوجيا المعلومات والاتصالات واستخدامها بالشكل الصحيح.
- توزيع دائرة الثقافة العلمية لتشمل الفيزياء والرياضيات والإحصاء.
- فرص عمل في مجالات مختلفة وبرواتب مرتفعة.

ثانياً: السلبيات

- الشعور بالملل أحياناً بسبب الأعمال الروتينية.
- التفكير العقلاني المستمر وبذل الكثير من الجهود العقلية.
- التعب والإرهاق بسبب مواصلة العمل على مدار ساعات طويلة.
- دراسة التخصص صعبة إلى حد ما.
- يحتاج التخصص إلى المتابعة المستمرة لآخر التطورات في المجال.
- عدم توافر برامج تدريس قوية للتخصص في الجامعات العربية.

- انعدام الخصوصية لصعوبة إخفاء هوية أصحاب البيانات أحياناً.
- تكلفة دراسة التخصص تعد عالية.

ومن أشهر الجامعات التي تقوم بتدريس علم البيانات:

- جامعة كارنيجي ميلون Carnegie Millon University.
- جامعة ستانفورد Stanford University.
- جامعة سانتا كلارا Santa Clara University.
- جامعة ميتشيغان University of Michigan.
- جامعة تكساس في دالاس University of Texas at Dallas.
- جامعة فيرجينيا University of Virginia.
- جامعة فلوريدا University of Florida.
- جامعة بوردو Purdue University.
- جامعة ميريلاند University of Maryland.
- جامعة جورجيا للتكنولوجيا Georgia University of Technology.
- جامعة الأميرة سمية Princess Sumaya University.
- جامعة روشترة University of Roche (U.S.A).
- جامعة نيويورك New York university.
- جامعة وارويك University of Warwick.

برنامج مايكروسوفت المهني في علم البيانات (Microsoft Professional Program for Data Science) :

تقدم شركة مايكروسوفت برنامجاً مهنياً متخصصاً في علم البيانات يضم 10 مساقات إلكترونية وتقدم بشكل إلكتروني من خلال مجموعة من الخبراء والأكاديميين ، وهى :

- مقدمة في علم البيانات.

- الاستعلام عن البيانات (باستخدام لغة البرمجة SQL).
- تحليل وتصوير البيانات باستخدام (Excel) ولغة (Power B I).
- الإحصاء لتحليل البيانات.
- استكشاف البيانات باستخدام الشيفرة (باستخدام لغة (R) ولغة (Python).
- المفاهيم الأساسية في علم البيانات (عمليات علوم البيانات) ولغة (Numpy).
- التعلم الذاتي للآلة. (نظرية التعلم العميق وتطبيقاتها).
- نمذجة البيانات (باستخدام الشيفرة أو الكود).
- علم البيانات التطبيقي.
- مشروع التخرج.

وتقدم شركة (IBM) الرائدة في صناعة وتطوير الحواسيب والبرمجيات شهادة مهنية في علم البيانات بالتعاون مع منصة (كورسيرا) للتعلم الإلكتروني، وتتكون الشهادة من 9 دورات تدريبية تتناول الموضوعات التالية:

1- ما هو علم البيانات؟
2- أدوات مفتوحة المصدر لعلم البيانات.
3- منهجية علم البيانات.
4- لغة باثيون (Python) لعلم البيانات والذكاء الاصطناعي.
5- قواعد البيانات ولغة الاستعلامات الهيكلية.
6- تحليل البيانات.
7- تصوير البيانات.
8- تعلم الآلة.
9- علم البيانات التطبيقي.

المراجع

الجعيد، عزيز عوض الله. من هو عالم البيانات وكيف يمكنني أن أصبح واحداً؟ aziz.blog.com تاريخ الدخول 2020/3/26م.

الخياط، إياد (2018). ما هو علم البيانات؟ datasciencearabi.com. تاريخ الدخول 2020/3/25م.

علم البيانات Data science Forqz.com. تاريخ الدخول 2020/3/25م.

علم البيانات. ويكبيديا. ar.m.wikipedia.org تاريخ الدخول 2020/3/25م

Stanton, Jeffery (2013). Introduction to data science.

Tolle, Kristin. (2009). The fourth paradigm: Data-intensive scientific. Microsoft Research.

What is data science? Data science. Bereley.edu. تاريخ الدخول 2020/3/28م.

What is data science and what does a data scientist do? www.kdnuggets.com. تاريخ الدخول 2018/9/7

Zhu, Yangyong (2015). Towards data science. Data Science Journal. (14). ISSN 1683-1470.

Data science: the science of the future

Rebhi Mustafa Alyan

Prof. of library and Informaition science

Faculty of Educational science – University of Jourdan

Abstract:

Data science is one of the new sciences that emerged at the beginning of the twenty-first century in developed countries, especially the United States of America, which teaches it at the bachelor's, master's and doctoral levels. The study deals with this new discipline, the data scientist, the pros and cons of studying data science, the most famous universities that teach it, and the Microsoft professional program in data science.

Keywords: Data science; Data scientists; Data analysts; data science Programs.