المنظمة العربية للتربية والثقافة والعلوم

تدعيم استخدام تحليلات التعلم والبيانات الضخمة في مجال التعليم في الوطن العربي

Learning Analytics & Big Data

مقدمة عامة 4

البيانات الضخمة 7

1. مقدمة 7

2. أسس البيانات الضخمة 8

3. تاريخ البيانات الضخمة 9

4. تعريف البيانات الضخمة 11

5. اهتمامات البيانات الضخمة 13

6. أسباب ظهور البيانات الضخمة 14

7. خصائص البيانات الضخمة 15

7. 1 البيانات الضخمة والحجم 16

7. 2 البيانات الضخمة والسرعة 20

7. 3 البيانات الضخمة والتنوع 22

7. 4 البيانات الضخمة والصدق 26

7. 5 البيانات الضخمة والقيمة 27

8. مصادر البيانات الضخمة 29

9. تقنيات البيانات الضخمة 29

10. الأطراف فى منظومة البيانات الضخمة 30

11. تطبيقات البيانات الضخمة 31

11. 1 القطاع الحكومى والبيانات الضخمة 31

11. 2 تحليل البيانات الضخمة و تحسين التعليم 32

11. 3 تحليل البيانات الضخمة لتحسين عملية صنع القرار 33

11. 4 الاستفادة من البيانات الضخمة في المجال العسكري 34

11. 5 الاستفادة من البيانات الضخمة في المجال الاقتصادي 34

12. الخاتمة 36

تحليلات البيانات 38
1. مقدمة 38
2. استراتيجية بناء منظومة تحليل البيانات 39
3. مصادر البيانات 39
4. أصناف البيانات 42
5. أنواع تحليلات البيانات 44
5. 1 تحليل النصوص 45
5. 2 تحليل الخطاب 47
5. 3 تحليل الصور والفيديوهات 47
6. عرض نتائج تحليل البيانات 49
6. 1 العرض المرئي للبيانات 49
6. 2 الأساليب المرئية الحديثة لعرض البيانات 52
7. تعلم الآلة 55
7. 1 نبذة حول المبادئ العامة لتعلم الآلة 56
7. 2 مراحل بناء نظم تعلم الآلة 57
7. 3 فئات خوارزميات تعلم الآلة 59
7. 4 تقييم خوارزميات تعلم الآلة 65
8. الخاتمة 70
تحليلات التعلم 71
1. مقدمة 71
2. المفاهيم الأساسية لتحليلات التعلم 73
2. 1 الدوافع 73
2. 2 التعريف بتحليلات التعلم 74
2. 3 مستخدمو تحليلات التعلم 75
2. 4 تحليلات التعلم و التنقيب عن البيانات التعليمية 79

2. 5 المستويات الأربعة لتحليلات التعلم 84
3. جمع البيانات التعليمية 89
3. 1 المقدمة 89
3. 2 جمع البيانات 89
3. 3 جمع البيانات في المساقات الضخمة المفتوحة عبر الإنترنت 90
3. 4 جمع البيانات في بيئة التعلم المعززة بالتكنولوجيا 92
3. 5 جمع بيانات القنوات المتعددة 94
3. 6 الأخلاقيات وخصوصية البيانات 96
4. أدوات تحليلات البيانات واستخداماتها 98
4. 1 لوحة القيادة 98
4. 2 العرض المرئي للبيانات في لوحات القيادة 102
4. 3 لوحة القيادة في MOODLE 105
4. 4 لوحة تحليلات يوتيوب 110
4. 5 تحليلات التعلم التنبؤية 119
5. الخاتمة 121
الخاتمة والمقترحات 122
المراجع 127

مقدمة عامة

لاريب أن تقدم الأمم كان ولا زال مرتبطا ارتباطا وثيقا بالتعليم كأداة أساسية للرقي بالمجتمعات وذلك على كل المستويات الثقافية منها والاجتماعية والاقتصادية. واليوم في عصر المعلومات والتكنولوجيات الحديثة أصبح التعليم

أكثر من أي وقت مضى ضرورة لا مناص عنها. وقد شهدنا, خاصة منذ بداية القرن الحادي والعشرين, تنافسا حادا بين الدول والأمم في مجال التعليم نتبينه من عدة مؤشرات منها تعدد التصنيفات وشدة الاهتمام بها لما تمثله من أداة تقييمية هامة لمجهودات النهوض بواقع التعليم وتحسين مخرجاته. من بين هذه التصنيفات على سبيل الذكر لا الحصر نجد كل من تصنيف شنغهاي للجامعات[1] وتصنيف World population review[2] وتصنيف Word top 20[3] وتصنيف QS World Ranking Universities[4]. وهو غيض من فيض من التصنيفات التي تعتمد كمؤشرات على التقدم التكنولوجي والاقتصادي بل والاجتماعي والحضاري للدول والمجتمعات. ولعل من أهم التحولات التي يعرفها مجال التعليم هو العناية المتزايدة بالكيف قبل الكم. إذ أن مختلف التصنيفات والتقييمات لمنظومات التعلم والتعليم تعتمد مؤشرات نوعية. بالتالي أصبحت مسألة تقييم جودة العملية التعليمية أساسية في النظم التعليمية الحديثة وذلك قصد اتخاذ الخطوات التصحيحية اللازمة عند الاقتضاء .

من ناحية أخرى أخذ التعليم الالكتروني حيزا هاما من المجال التعليمي لما له من ميزات من ناحية, ومن ناحية أخرى لما عرفه من نضوج على مستوى الأدوات والمقاربات والاستعمالات. كما أن انتشار وباء COVID-19 منذ سنة 2020 مثل عاملا هاما عزز من مكانة التعليم الالكتروني. حيث شكلت سياسات الإغلاق حافزا لإعطاء دفعة هامة للتعليم الإلكتروني. إذ مثل اعتماده ضمانة لاستمرارية التعليم خلال ذروة الجائحة. في هذا الإطار وجبت الاشارة لما يمثله هذا النمط التعليمي من فرصة لجمع البيانات حول العملية التعليمية. فمن المعلوم

[1] https://www.shanghairanking.com

[2] https://worldpopulationreview.com/country-rankings/education-rankings-by-country

[3] https://worldtop20.org/worldbesteducationsystem

[4] https://www.qschina.cn/en/university-rankings/university-subject-rankings/2021/education-training

أن النظم المعلوماتية تمثل منجم من المعطيات والبيانات التي تشمل مختلف نواحي استعمالها. كما أن جمع البيانات الرقمية ومعالجتها وتخزينها والتصرف فيها أفضل وأسرع من معالجة البيانات الغير رقمية.

جانب المعالجة والتخزين تحديدا عرف تحديات جديدة مع الطفرة الكبيرة التي عرفها استعمال الأنظمة المعلوماتية عموما ومنها أنظمة التعليم الإلكتروني. في هذا الإطار يتنزل مفهوم البيانات الضخمة وما صاحبه من أعمال بحث وتجديد. الهدف منها هو توفير خوارزميات وبنى تحتية معلوماتية وحلول تكنولوجية تسمح بالتعامل مع ضخامة البيانات التي تتجاوز القدرات البشرية للتعامل معها. بل وتتجاوز حتى قدرات المنظومات المعلوماتية العادية.

البيانات عموما والبيانات الضخمة خصوصا تحتوي جملة من المعطيات والمعارف الهامة. الإشكالية هنا تتمثل في القدرة على اكتشافها وتبويبها وحسن عرضها لاستغلالها واستثمارها. هنا تأتي أهمية علم تحليل البيانات الذي شهد تطورا كبيرا مع التقدم الكبير الذي عرفه الذكاء الاصطناعي وخاصة الطفرة التي عرفتها خوارزميات تعلم الآلة. في هذا الإطار يعتبر تحليل البيانات (Data analytics) وسيلة تسمح بتحليل وبقياس البيانات بهدف أخذ أفضل القرارات وتحسين المردودية. أما مجالات تطبيقه فعديدة, منها المجال الصناعي, والمجال التجاري والمجال الطبي على سبيل الذكر لا الحصر.

يعتبر مجال التعليم أيضا من جملة المجالات التي يمكنها التعويل على تحليل البيانات من خلال مفهوم تحليلات البيانات التعليمية (Learning analytics). إذ أن الانتشار الواسع للتعليم الإلكتروني من جهة ووفرة بيانات التعلم وضخامتها من جهة أخرى سمحت بتطبيق معارف وتقنيات تحليل البيانات على المجال التعليمي. أما الهدف من هذه التحليلات بصفة عامة فيتمثل في تعزيز الجودة والتنافسية, والتعرف على أفضل الممارسات (Best

practices), والحوكمة الرشيدة, والشفافية, والتحسين المستمر.

هذه الدراسة تتوجه لكل من أصحاب القرار في المجال التعليمي و المشرفين على النظم التعليمية و للأساتذة و المعلمين بهدف تقديم نظرة شاملة عن مفاهيم البيانات الضخمة و تحليل البيانات و تحليلات بيانات التعليم خاصة. إذ تراوح الدراسة بين تقديم المفاهيم النظرية وتطبيقاتها بهدف تكوين نظرة شاملة تسمح بالاستفادة منها وتبنيها من أجل مستقبل للتعليم في العالم العربي. كما ننوه أن القارئ لا يحتاج لأي مؤهلات أو معارف أو مهارات خاصة من أجل الاستفادة.

هذه الدراسة تضم ثلاثة أبواب. الباب الأول يتناول بالدرس مفهوم البيانات الضخمة. أما الباب الثاني فإنه يعنَى بمفهوم تحليل البيانات. أما الباب الثالث فيتعلق بمفهوم تحليلات التعلم. في ختام الدراسة قمنا بإدراج جملة من المقترحات والتوصيات ذات العلاقة.

والله ولي التوفيق

البيانات الضخمة

1. مقدمة

البيانات الضخمة، الجميع يتحدث عنها ومع ذلك، لا ندرك مدى تأثير هذه التقنيات الجديدة علينا كل يوم. بدون أن ندرك ذلك، أصبحت حياتنا "رقمية" أكثر فأكثر وتولد تدفقًا لا نهاية له من البيانات. هل هذا الطوفان من البيانات يُحدث ثورة في الطريقة التي نختبر بها العالم ونفهمه، وحتى ... نتنبأ به؟ من هناك تبرز عديد من الاسئلة:

- ما هي البيانات الضخمة؟
- هل يمكننا توقع العالم بالبيانات الضخمة؟
- لماذا نقول إن العالم أصبح خوارزميًا؟
- هل هناك خطر من إساءة استخدام خصوصيتنا؟
- ما هي مجالات تطبيق البيانات الضخمة؟
- ما هي تجارب البيانات الضخمة في مجال التعليم؟

هذا ما سنحاول اكتشافه في هذا الباب من خلال العديد من الأمثلة الحقيقية التي يبدو أنها تأتي مباشرة من أفلام الخيال العلمي.

2. أسس البيانات الضخمة

نقطة البداية: أفلام هوليوود. كيف استطاعت هذه الأفلام أن تتوقع كل التطورات التكنولوجية التي نشهدها اليوم؟ لماذا اختيار أفلام هوليوود المشهورة جدًا لشرح البيانات الضخمة[5] ؟ نشر المؤلفون Xavier Perret و Guy Jaquemelle و Stephane Richard في عام 2014 كتابًا بعنوان "البيانات الضخمة: السينما قد خططت بالفعل لكل شيء" شرحوا فيه من خلال أمثلة توضيحية كيف ظهرت البيانات الضخمة لفترة طويلة في هوليوود. فيما يلي نقدم الأمثلة التي اقترحوها.

فيلم 1: ليتل بيج مان – فيلم سياسي – 1970 – الممثل الرئيسي: داستن هوفمان – إنه فيلم غربي يتحدث عن الملحمة الهندية ويحكي قصة حرب فيتنام.

فيلم 2: ليلة الموتى – فيلم سياسي – 1968– الفيلم يحكي الخوف من الطاقة النووية والكوارث المحتملة.

فيلم 3: 2001 رحلة فضائية – فيلم روائي – 1968 – يحكي الفيلم قصة رواد الفضاء عندما واجهوا كتلة متراصة سوداء غامضة سقطت من السماء. في هذا الفيلم ، كان هناك العديد من الاختراعات التكنولوجية التي تعتمد بالفعل على الذكاء الاصطناعي ، مع الكمبيوتر القوي Hal (تدوين IBM). تسبب هذا الفيلم في حدوث صراع بين Samsung و Apple لأن أحد المشاهد أظهر كائنًا في عام 1968 يشبه الكمبيوتر اللوحي. لذلك، منذ عام 1968، كنا نشهد بدايات التقنيات التي نشهدها اليوم.

فيلم 4: تقرير الأقلية (Minority Reports) – فيلم خيالي – 2002 – الممثل الرئيسي: توم كروز– هذا الممثل الذي لعب بأشياء افتراضية ، ودخل سيارة أجرة بدون سائق ، ومن هنا جاءت مباني المركبات المستقلة التي نراها اليوم.

---

5 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

السؤال الذي يطرح نفسه : كيف استطاع هؤلاء المخرجون توقع المستقبل؟

في عام 1968 ، أحاط ستانلي كوبريك نفسه بأفضل الباحثين والعلماء في العالم ، لمناقشة ما يجب أن يكون عليه المستقبل ، وكان أيضًا عصر ظهور الذكاء الاصطناعي من خلال الباحث آلان تورينج .[6]

نفس الشيء بالنسبة لمدير Minority Reports ، ستيفن سبيلبرغ ، ولمدة 3 سنوات ناشد جميع علماء الكوكب لإعادة إنشاء مستقبل يجب أن يكون معقولًا قدر الإمكان. المثير في هذا الفيلم هو أن المشاهد تعرض مواقف ندركها اليوم بفضل البيانات الضخمة.

فيلم آخر ليس بعيدًا : فيلم Interstellar لعام 2014، والذي يتنبأ بالمستقبل بعد مرور 30 أو 40 عامًا. المثير في هذا الفيلم هو أن المخرج كريستوف نولان فعل نفس الشيء مثل الآخرين، وأحاط نفسه بالعديد من العلماء لتوقع التطورات التكنولوجية .[7]

السؤال ماذا تفعل البيانات الضخمة في كل هذا؟

في الواقع، في كل هذه الأفلام، يتم استخدام قدرًا كبيرًا من المعلومات بأشكال مختلفة، لذلك كانت البيانات الضخمة موجودة بالفعل منذ خمسين عامًا مع تطبيق الخوارزميات عليها أيضًا.

3. تاريخ البيانات الضخمة

البيانات الضخمة موجودة منذ وقت طويل. ربما بدأت في القرن الثالث قبل الميلاد, عندما قرر أحد جنرالات الإسكندر الأكبر، بناء مكتبة الإسكندرية الشهيرة، من أجل جمع كل معارف العالم في ذلك الوقت. بسرعة كبيرة،

6 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

7 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

تم إدراج 50000 عمل في هذه المكتبة. بعد ثلاثة قرون وصلنا إلى 700000 عمل. اليوم، إذا قسمنا كمية المعلومات المتوفرة على الأرض على عدد السكان، فلكل واحد نصيب يبلغ 320 ضعف الكمية المتوفرة في هذه المكتبة .[8]

لقرون وقرون، كان يُطلب من الناسخين نسخ كل هذه المعلومات. في عام 1454 جاء جوتنبرج بتقنية الطباعة وقام بطبع إنجيل جوتنبرج الشهير. حيث تم طبع 180 نسخة.

الفترة ما بين 1454-1504 كانت 50 سنة غير عادية من حيث إنتاج البيانات المطبوعة. إذ طبع ما يقارب 8 ملايين كتاب خلال هذه السنوات الخمسين.

قد أدى كل ذلك إلى ثورة في الوصول إلى المعلومات ومشاركتها. بالنظر إلى القرن العشرين والنظر في بعض التواريخ المهمة، اتضح ما يلي:

- في عام 1941، تحدث قاموس أوكسفورد الإنجليزي عن انفجار البيانات. هذا تغيير جذري لأن في ذلك الحين كانت المعلومات لا تزال نادرة وثمينة.
- بعد 3 سنوات، في عام 1944، قام الأكاديمي فريمان تردر بحساب عدد الكتب في المكتبات واستنتج أن العدد يتضاعف كل 16 عامًا. بحلول عام 2040، سيكون هناك 200 مليون نسخة.
- في عام 1950 كان هناك الأعمال والبحوث العلمية لعالم الرياضيات الشهير آلان تورينج.
- بعد 10 سنوات، عمل الباحثون على مفهوم الأسي في الرياضيات لشرح نمو المعلومات، وخاصة على وسائل وآليات ضغطها.
- في عام 1968، ظهر فيلم "2001: A Space Odyssey" بحاسوب HAL العملاق الذي جمع كمية كبيرة من المعلومات.

---

8 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

- في عام 1971، نشر آرثر مولر كتابًا عن حماية البيانات والخصوصية. هذه القضية لا تزال ذات صلة اليوم.
- 1975-1976: ظهور الميكرومعلوماتية مع وصول شركة IBM وApple.
- في عام 1980، كان قانون باركنسون: سعة التخزين تنمو بشكل كبير وكذلك تطور المعلومات.
- في عام 1983، ظهر فيلم "Dead Zone". القصة تحكي عن شخص تعرض لحادث سيارة. في ذهنه يمتزج الماضي والحاضر والمستقبل معًا. السؤال هو، هل يمكننا تغيير المستقبل بمجرد توقعه.
- في التسعينيات، كانت هناك ثورة الإنترنت: تتدفق المعلومات بشكل أسرع وأسرع، وتتغير بسرعة كبيرة. والبيانات الضخمة في كل هذا؟ تم استخدام مصطلح البيانات الضخمة لأول مرة في عام 1997 من قبل اثنين من باحثي ناسا .[9]
- بعد عامين في عام 1999، إصدار فيلم Matrix: لأول مرة نرى طوفانًا من البيانات على الشاشة.
- في عام 2008، احتلت BIG DATA الصفحة الأولى في المجلة المرموقة "Nature".
- في عام 2013، صدر فيلم "Her"[10]: الممثل سيقع في حب نظام التشغيل الخاص به .

4. تعريف البيانات الضخمة

البيانات الضخمة تشير إلى مجموعة كبيرة جدًا من البيانات والخصائص الأخرى التي لا يمكن لأداة إدارة قاعدة البيانات التقليدية (Data Base Management System) أو إدارة المعلومات (Information Management System) التعامل معها تمامًا. لشرح هذا المفهوم وإعطاء أرقام عن البيانات الضخمة ،

9 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

10 Big Data : Le Cinéma avait déjà tout prévu, de Xavier Perret, Guy Jaquemelle, Stephane Richard, Edition Kawa 2014

اعتمدنا على الدورة التي قدمتها جامعة IBM حول تحليلات البيانات الضخمة ، في عام 2017[11] وكتاب المؤلف برنارد مار بعنوان "البيانات الضخمة في الممارسة: كيف استخدمت 45 شركة ناجحة تحليلات البيانات الكبيرة لتقديم نتائج استثنائية " والذي نُشر في عام 2016. [12]

إذا, تمثل البيانات الضخمة مرحلة هامة من مراحل تطور نظم المعلومات والإتصالات. وهي تعبر في مفهومها المبسط عن كمية هائلة من البيانات المعقدة التي يفوق حجمها قدرة البرمجيات والآليات الحاسوبية التقليدية على تخزينها ومعالجتها وتوزيعها. هذا الأمر أدى إلى وضع حلول بديلة متطورة تمكن من التحكم في تدفق البيانات والسيطرة عليها.

تمتلك تقنية البيانات الضخمة إمكانية تحليل بيانات مواقع الإنترنت وأجهزة الإستشعار وبيانات شبكات التواصل الإجتماعي، حيث أن تحليل هذه البيانات يسمح بوجود ارتباطات بين مجموعة من البيانات المستقلة لكشف جوانب عديدة. من هذه الجوانب على سبيل المثال التنبؤ بالتوجهات التجارية للشركات ومُكافحة الجريمة فى المجال الأمني وغيرها. كما توفر هذه التنبؤات لصانعي القرار أدوات مبتكرة لفهم أفضل للظروف وبالتالي اتخاذ قرارات صحيحة تحقق الأهداف المطلوبة.

في الواقع، نحن ننتج 2.5 تريليون بايت (1 تريليون = 10-18 => 2.5 تريليون) بايت من البيانات كل يوم. هذه البيانات متنوعة المصادر: الرسائل التي نرسلها، والرسائل من وسائل التواصل الاجتماعي، ومقاطع الفيديو الرقمية والصور التي ننشرها، والمعلومات المناخية المستغلة من أجهزة الاستشعار، وإشارات GPS من الهواتف

11 IBM,BigData-University, "BigDataAnalytics – Lesson1: What is BigData", 2017.

12 Marr, B., "BigData:The 5Vs Everyone MustKnow", from the Book "Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results", Hardcover, 2016

المحمولة، وسجلات المعاملات للتسوق عبر الإنترنت وغيرها الكثير. 90٪ من البيانات المنتجة غير منظمة. تسمى هذه البيانات، البيانات الضخمة.

عمالقة الويب، وفي مقدمتهم Facebook وGoogle وYahoo، كانوا أول من نشر هذا النوع من التكنولوجيا والأدوات اللازمة للتعامل معها وتحليلها.

إن مفهوم البيانات الضخمة هو مفهوم شاع في عام 2012 ليعكس حقيقة أن الشركات تواجه كميات كبيرة من البيانات وبشكل متزايد والتي يتعين معالجتها. والهدف هو الحصول على حصة تجارية وتسويقية أكبر. غير أنه من الصعب استخدام هذه البيانات الضخمة مع أدوات إدارة قواعد البيانات الكلاسيكية.

لذلك فالبيانات الضخمة هي عبارة عن مجموعة من التقنيات، والهندسة، والأدوات والإجراءات التي تسمح للمؤسسة بالتقاط ومعالجة وتحليل كميات كبيرة ومحتوى غير متجانس ومتغير بسرعة كبيرة، واستخراج المعلومات منها ذات الصلة, بتكلفة معقولة.

5. اهتمامات البيانات الضخمة

مثلما أشرنا، فإن البيانات الضخمة التي تأتي من مصادر متعددة (مواقع، وقواعد بيانات، وهواتف، وخوادم)، ستكون ذات فوائد كبيرة [13] لـ:

- الكشف عن مشاعر وردود فعل العملاء.
- الكشف عن الحالات الحرجة أو التي تهدد الحياة في المستشفيات وفي الوقت المناسب للتدخل.
- توقع نماذج الطقس لتخطيط الاستخدام الأمثل لتوربينات الرياح.

13 IBM,BigData-University, "BigData Analytics – Lesson1:What is BigData", 2017.

– اتخاذ قرارات محفوفة بالمخاطر بناءً على بيانات المعاملات في الوقت الفعلي.

– التعرف على المجرمين والتهديدات من مقاطع الفيديو والأصوات وتدفق البيانات.

– دراسة ردود أفعال الطلبة خلال الدورة، وتوقع من سينجحون، وفقًا للإحصاءات والنماذج التي تم جمعها على مر السنين (مجال البيانات الضخمة في التعليم).

هذا على سبيل الذكر, حيث أن مجالات استخدام البيانات الضخمة غير محدود. هاته الاهتمامات تجعل الباحثين والصناعيين في المجال التكنولوجي أمام العديد من التحديات، نذكر منها بالأساس:

– جمع كمية كبيرة من البيانات المتنوعة للعثور على أفكار جديدة.

– التقاط البيانات التي تم إنشاؤها بسرعة.

– حفظ كل هذه البيانات بطريقة ذكية.

– معالجة هذه البيانات واستخدامها لصنع القرار أو لأغراض استراتيجية.

6. أسباب ظهور البيانات الضخمة

في الواقع، ظهرت البيانات الضخمة نتيجة للجمع بين 3 ظواهر: قوة المعالجات الدقيقة، وسعة التخزين وسرعة الشبكات. [14]

– بالنسبة لقوة المعالجات الدقيقة فقد تم مضاعفة السعة بمقدار مليون في 40 سنة، زد على هذا، تتمتع الهواتف الذكية اليوم بأداء لا يضاهى مقارنة بأجهزة كمبيوتر كبيرة من السبعينيات.

14 IBM,BigData-University, "BigData Analytics– Lesson1: What is BigData", 2017.

- بالنسبة للتخزين فقد انخفضت تكاليفه بصفة كبيرة: 1 جيجابايت كانت تكلف مليون دولار منذ 40 عاماً، أما اليوم تكلف 1 غيغابايت أقل من دولار واحد.
- بالنسبة لسرعة الشبكة فمنذ 20 عامًا، كان المودم يعمل بسرعة 9600 ب/ث، أما اليوم فتوجد اليوم أجهزة مودم تصل سرعتها إلى 10 جيجابيت / ثانية وما فوق.

لإدارة البيانات الضخمة، سيكون من الضروري: جمعها، تخزينها، تحليلها واتخاذ قرارات بعد هذا التحليل (توقع الحوادث، عمل التنبؤات، إلخ).

7. خصائص البيانات الضخمة

لشرح خصائص البيانات الضخمة ، اعتمدنا على الدورة التي قدمتها جامعة IBM حول تحليلات البيانات الضخمة ، في عام 2017 وكتاب المؤلف برنارد مار بعنوان "البيانات الضخمة في الممارسة: كيف استخدمت 45 شركة ناجحة تحليلات البيانات الكبيرة لتقديم نتائج استثنائية – الجزء : خصائص Vs5 في BigData التي يجب على الجميع معرفتها" والذي نُشر في عام 2016. وكذلك على مقال أعد من طرف عدنان مصطفى البار وخالد علي مرحابي (جامعة الأمير عبد العزيز وجامعة أم القرى – المملكة العربية السعودية) بعنوان "البيانات الضخمة ومجالات تطبيقها" الذي صدر في عام 2018.

تتميز البيانات الضخمة بخمس خصائص تسمى Vs5 وهي: السرعة (Velocity)، الحجم (Volume)، التنوع (Variety)، الصدق (Veracity) و القيمة (Value). لم يتم تقديم هاته الخصائص في وقت واحد، لقد بدأنا بـ 3، ثم 5، واليوم يمكن تعداد أكثر من 10 خصائص. فيما يلي، سنقدم هذا التطور بمرور الوقت:

- في عام 2001, 3 خصائص، حددتها Gartner: الحجم، التنوع والسرعة.

- في عام 2012, 4 خصائص: أضافت IBM العنصر الرابع وهو Veracity أي الصدق.
- في عام 2014, ظهرت الخاصية الخامسة : وهي القيمة.
- في عام 2016, 7 خصائص: تمت إضافة اثنين التباينٍ (Variability) والتصور (Visualization).
- في عام 2017, 10 خصائص, حيث تمت إضافة الجدوى (Virality) والحيوية (Viability) و اللزوجة (Viscosity).

بشكل عام ، من يقول البيانات الضخمة يقول V5: الحجم / السرعة / التنوع / الصدق / القيمة. في هذا القسم سوف نتناول بالدرس أولى هاته الخصائص ، أي الحجم.

7. 1 البيانات الضخمة والحجم

7. 1. 1 المشكلة

انخفض سعر تخزين البيانات كثيرًا في الثلاثين عامًا الماضية: من 100000 دولار / جيجابايت (1980) إلى 0.10 دولار / جيجابايت (2013) و 0.02 / جيجابايت (2021). لكن في الواقع ، يمكن أن تكون مواقع التخزين الموثوقة (مثل شبكات منطقة التخزين SAN: Storage Area Network) أو شبكات التخزين عالية الكلفة.

لذلك ، كان من الضروري اختيار تخزين بعض البيانات فقط ، التي تعتبر حساسة ، ومن هنا تأتي مخاطر فقدان البيانات ، والتي قد تكون مفيدة للغاية ، مثل السجلات Logs على سبيل المثال.

السؤال الأول إذا هو كيفية تحديد البيانات التي تستحق التخزين؟ (البيانات من المعاملات؟ البيانات من ملفات السجل؟ البيانات التجارية؟ بيانات المستخدم؟ استرجاع البيانات من أجهزة الاستشعار؟ البيانات الطبية؟ البيانات الاجتماعية؟)

السؤال الثاني : كيف يتم تخزين البيانات في مكان موثوق به أقل تكلفة؟

السؤال الثالث : كيف يتم تصفح هذه البيانات واستخراج المعلومات بسهولة وبسرعة؟

2.7.1 الحجم (**Volume**)

الرقمنة هي سبب الانتشار الغير مسبوق للبيانات والبيانات الضخمة. إذ حتى عام 2000 ، كانت 20٪ من البيانات فقط رقمية. ثم في العام 2015 ، أصبح 98٪ من البيانات رقميًا. وتقدر نسبة البيانات الرقمية المخزنة من جملة البيانات المخزنة ب 99٪ في 2020.

ما هي الأدوات التي يمكن أن توفر هذه البيانات؟

أولاً ، يجب أن يكون العالم متصلاً (وهو كذلك).

- في عام 2015: بيع 2.5 مليار محطة طرفيةTerminal)) أي ما يقارب ثلث سكان العالم.
- في عام 2020: 4.25 مليار محطة طرفية (Terminal) من أصل 7.8 مليار نسمة (أكثر من 50٪).

ثانيا يجب مراكمة البيانات (وهو ما يحدث فعلا).

شركتان ترمزان إلى تراكم البيانات هما Google و Facebook. بذلت هاتان الشركتان كل شيء للاحتفاظ بالمستخدمين من أجل تجميع أكبر قدر ممكن من البيانات لتحقيق الدخل من المعلنين. وقد إستدعى هذا من

الشركتين الاحتفاظ بكميات ضخمة جدا من البيانات. حيث تم إحصاء في كل دقيقة من سنة 2021 الكم التالي من البيانات الجديدة المنشورة من قبل المستخدمين:

- 700000 صورة على الانستغرام
- 450000 تغريدة
- 70.000.000 رسالة فورية
- 220 مليون بريد إلكتروني

أصبحت Google ضرورية: في أوروبا ، تشير التقديرات إلى أن حصتها في السوق من حيث البحث تزيد عن 90٪. الموقع الذي لم تتم الإشارة إليه بواسطة Google لديه فرصة منخفضة للبقاء على قيد الوجود. ولعل الأرقام التالية تغني عن كل تعليق:

- في عام 2016: 5.5 مليار عملية بحث يوميًا.
- في عام 2021: 7 مليارات عملية بحث يوميًا.
- مع YouTube على Google ، هناك 2 مليار مستخدم في عام 2021.
- مع Android ، هناك 2.5 مليار مستخدم في عام 2021.
- في عام 2013 ، تسبب انقطاع خدمة Google لمدة 5 دقائق في انخفاض حركة الإنترنت العالمية بنسبة 40٪.

تهتم Google أيضًا بالعالم الحقيقي حيث قامت برقمنة أكثر من 15 مليون كتاب. كل ما هو غير رقمي يصبح رقمي مع Google. وفي هذا الإطار أنشأت Google خدمة تتيح تحليل تكرار الكلمات على مر

السنين والقرون وهي Google Ngrams[15] اعتمادا على محتوى الوثائق المرقمنة. مثال: حالة الأنفلونزا. يُظهر استخدام كلمة إنفلونزا بمرور الوقت أن ذروة الاستخدام تحدث عند ظهور وباء أو مرض إنفلونزا. يمكننا أن نسلك نفس المنهج لتحليل جائحة مرض فيروس كورونا.

شركة ضخمة أخرى أثرت بشكل مباشر في واقع البيانات الضخمة هي شركة Facebook التي تحولت إلى Meta Platforms, Inc. انطلق الأمر في عام 2003 حيث تخيل طالب شاب (مارك زوكربيرج) في غرفته شبكة اجتماعية. بعد 12 عامًا ، أصبح لدى Facebook مليارو400 ألف مستخدم. ليتطور الأمر في عام 2020، ليبلغ عدد المستخدمين 2.74 مليار مستخدم. وبلغ عدد المستخدمين 2.912 مليار في الربع الأول من 2022.

كما تجدر الإشارة إلى أنه في العام 2014 اشترى Facebook تطبيق whatsapp الذي يضم أكثر من 2 مليار مستخدم . كما تم الاستحواذ في عام 2012 على تطبيق Instagram الذي يضم أكثر من مليار مستخدم. بالتالي سمحت كل هذه التطبيقات لشركة Facebook أن تتحصل على مصادر متعددة وثرية للبيانات الضخمة.

Google و Facebook ليسا الوحيدين اللذين يقومان بفهرسة وإنشاء ونشر البيانات الرقمية. بالتالي السؤال المطروح هو ما حجم البيانات التي ننتجها إجمالاً من جميع المصادر والتطبيقات؟

في عام 2010 ، نشر إريك سميتش دراسة قام على أساسها بحساب المعلومات التي تم إنتاجها منذ بداية البشرية. في عام 2003 ، كان الرقم 5 إكسابايت مع العلم أن 1 إكسابايت = 10^{18} بايت = 1000 مليار ميغا

15 https://books.google.com/ngrams

بايت. بينما في عام 2010 ، استغرق الأمر يومين فقط لإنشاء هذه الـ 5 إكسابايت. في عام 2013 ، استغرق الأمر 10 دقائق فقط.

تجدر الإشارة هنا إلى أن هذه الغزارة في البيانات لا تعود فقط إلى استخدامنا كبشر للتطبيقات والمنصات الرقمية,بل تعود أيضا لظهور ما يسمى بالأشياء المتصلة (Connected things) التي تنتج بدورها الكثير من البيانات. والمقصود بالأشياء المتصلة هي مختلف الأجهزة القادرة على الاتصال بشبكة الإنترنت. من ضمنها يمكن أن نذكر الساعات الذكية، السيارات الذكية، المنازل الذكية, التجهيزات المنزلية الذكية, وغيرها.

لفهم حجم تأثيرها نسوق هنا جملة من الأرقام للدلالة على مدى تأثيرها الحاضر وفي المستقبل على جمع البيانات الضخمة. ففي العام 2020 ، تم إحصاء أكثر من 40 مليار من الأشياء المتصلة آتية من أماكن مختلفة (شركات ، أفراد ، أماكن عامة ، إلخ).

لعل أهم ما يجب أن ننتبه إليه هو أن اليوم ، كل شيء نقرأه ، نستمع إليه ، على الأجهزة المختلفة (الهواتف الذكية ، أجهزة الكمبيوتر ، الأجهزة اللوحية ، إلخ) ، يولد البيانات. فنحن نعيش في محيط من البيانات الرقمية تشير الأرقام التي ذكرناها إلى حجمه الكبير. هذا فيما يخص الحجم, فماذا عن سرعة المعالجة؟

7. 2البيانات الضخمة والسرعة

نحن ننتج كميات كبيرة بشكل متزايد من البيانات. ولكن ، أصبح من الضروري معالجتها بسرعة كبيرة (Velocity). اذا، يجب تحليل جميع البيانات التي نجمعها في أسرع وقت ممكن ، حتى نتمكن من توقع احتياجات الحريف، على سبيل المثال، كما تفعل NetFlix و Amazon وأخذ القرارات الواجبة في الوقت المناسب.

عندما يتعلق الأمر بتوقع مشاكل الصحة العامة (كوفيد19 على سبيل المثال) ، فمن الضروري تخزين هذه البيانات وتحليلها بسرعة عالية جدًا ، لأن أدنى تأخير قد يضعنا في مواقف مأساوية. تحتاج أيضًا إلى الحصول على معلومات أمام عينيك لاتخاذ القرارات الصحيحة. لذا فإن السرعة والتخيل هما الجانبان اللذان سيتم تفصيلهما في هذا القسم.

لتوضيح ما قلناه ،دعنا نتحدث قليلاً عن فيلم "Terminator" للمخرج جيمس كاميرون والذي لعب دوره الرئيسي أرنولد شوارزنيجر والذي صدر عام 1984. في هذا الفيلم ، كان أرنولد قد وضع نظارات مشابهة ل Google Glass (قبل 30 عامًا من تسويقها).

في Terminator 2 ، وفي مشهد مشهور في الحانة وبهذا النطاق ، يجب عليه مسح جميع حرفاء هذه الحانة. الهدف هو تحليل وتصور البيانات التي لدينا في الوقت الفعلي لحل مشكلة عاجلة (في مثالنا ، الضرورة الملحة هي العثور على المجرم).

أصبحت السرعة عنصرًا أساسيًا في طريقة فهم العالم ، وخاصة عندما يتعلق الأمر بالعالم الرقمي. ففي حالة محرك البحث جوجل ،يجب الإستجابة لـ 7 مليارات عملية بحث يوميًا. نرى أننا نحصل على استجابة فورية تقريبًا. فمن المهم من أجل تجربة مستخدم جيدة تقليل وقت الاستجابة بشكل أكبر ، وتقليل وقت إدخال استعلام المستخدم. حيث قدرت Google أن متوسط الوقت الذي يقضيه المستخدم في كتابة استعلام هو 9 ثوانٍ. وبمجرد أن يلاحظ Google الأحرف الأولى من الاستعلام ، فإنه يبدأ في الرد ؛ مما يسمح له بأخذ أسبقية ب ثانيتين.

هناك أمثلة أخرى حيث تعتبر السرعة عاملاً رئيسيا كبرنامج Watson الخاص بشركة IBM. وهو أحد برامج الذكاء الاصطناعي. وقد لوحظ في عام 2011 عندما فاز في لعبة Jeopardy الشهيرة. لهذا ، واجه

Watson أفضل المرشحين في العالم. للرد والتغلب على خصومه ، كان عليه تحليل 200 مليون صفحة في 3 ثوان.

المجال الطبي أيضا يزخر بأمثلة عديدة عن أهمية السرعة في تحليل ومعالجة البيانات الضخمة. فمثلا يعمل الدكتور ماكجريجور وفريقه في كندا على الأطفال الخدج. من خلال جمع وتحليل 1200 معلومة في الثانية (معدل ضربات القلب، ودرجة الحرارة، وما إلى ذلك) ، يمكنهم اتخاذ القرارات وتحليل معدل تدهور الحالة الصحية للطفل. مع مثل هذا النظام ، تمكنوا من توقع حالة الطفل قبل 24 ساعة ، مما سمح له بوصف الدواء في الوقت المناسب. بالتالي كانوا قادرين على إنقاذ حياة العديد من الأطفال.

إذا من المهم تحليل جميع البيانات التي لدينا في وقت وجيز لاتخاذ القرارات الصحيحة في الوقت المناسب. أي من المهم الحفاظ على خاصية السرعة (Velocity) عندما يتعلق الأمر بالبيانات الضخمة.

7. 3 البيانات الضخمة والتنوع

للتخزين في قواعد البيانات أو مستودعات البيانات ، يجب أن تحترم البيانات مسبقًا تنسيقًا محددًا. معظم البيانات الموجودة غير منظمة أو شبه منظمة. مع هذه المجموعة المتنوعة من البيانات (Variety)، نريد تخزين كل شيء. فعلى سبيل المثال بالنسبة للدردشة في مركز الاتصال ، يمكن تخزينها في شكل نصي لحفظ محتواها. كما يمكن تخزين التسجيل الصوتي كما هو للحفاظ على نبرة صوت الحريف مما يسمح بمعرفة مدى الرضا أو الغضب.

لذلك ، بعد الحجم والسرعة سنتحدث في هذا القسم عن تنوع البيانات. ترومان بطل الفيلم "ترومان شو" مثال ممتاز لتجسيد هذه الاستعارة. ولد ترومان في استوديو ومنذ ولادته كان بطل عرض الواقعي. يتابع ملايين

المشاهدين مغامراته في الوقت الفعلي: في أي وقت يستيقظ ، وماذا يفعل في الداخل ، وما يفعله في الخارج عندما يكون في العمل ، والعلامة التجارية لمعجون أسنانه. يعرف جميع المشاهدين كل شيء عن حياة ترومان.

تم الآن استبدال هؤلاء المشاهدين بمواقع الويب وتطبيقات الجوال التي تراقب سلوكنا باستمرار. بالتالي أصبح كل واحد منا هو نموذج لشخصية ترومان المذكورة في الفيلم.

اليوم ، ننتج مجموعة كبيرة ومتنوعة من البيانات ، بناءً على عمليات البحث التي نقوم بها و شبكاتنا الاجتماعية وساعاتنا الذكية. في الواقع ، هذه البيانات غير متجانسة (نصوص ، صور ، معدل ضربات القلب ، آراء على الشبكات الاجتماعية ، أماكننا المفضلة ، إلخ). كل من هذه البيانات هي جزء من اللغز. من خلال تجميع كل هذه الأجزاء معًا ، يمكن فهم كل واحد منا. مع كل هذا التنوع في البيانات ، يطرح السؤال حول نوعية قاعدة البيانات التي ستتعامل مع كل هذا.

قبل البيانات الضخمة ، كانت الوسيلة الأساسية لتخزين البيانات هي قواعد البيانات المنظمة. حيث يتم تحديد نموذج التخزين مسبقا وبصفة دقيقة, مع حد أدنى من المرونة. بينما عندما يتعلق الأمر بالبيانات الضخمة ، يصبح من المهم التعامل مع البيانات بطريقة مرنة تناسب طبيعتها المتغيرة.

باستخدام البيانات الضخمة ، يتعين علينا إدارة قواعد البيانات غير المهيكلة ، والتي يمكن إضافة أعمدة وصفوف جديدة إليها في أي وقت. وهكذا نكون قادرين على إدارتها بشكل ديناميكي للغاية.

في الواقع ، نحن كأفراد ننتج نوعين من البيانات: البيانات التعريفية: كالاسم والعمر والبيانات السلوكية: الموسيقى التي نستمع إليها ، والوقت الذي نستيقظ فيه ، والطرق التي نتخذها كل يوم ، والعلامة التجارية لمعجون الأسنان الخاص بنا.

هذه البيانات تهمنا ولكنها تهم أيضًا شركات الإعلان وشركات التأمين والبنوك ومواقع التجارة الإلكترونية.

نقطة أخرى يجب التنبه لها, وهي أن البيانات تنقسم إلى بيانات حقيقية وبيانات مشتقة. هذا يزيد من غزارة وتنوع البيانات. لتبيان الفرق بين هذين القسمين لنأخذ كمثال تطبيقة Facebook.

البيانات الفعلية هي البيانات الناتجة عن النقر على زر الإعجاب. أما البيانات المشتقة هي البيانات التي تم الحصول عليها بعد تنفيذ الخوارزميات القادرة على تحليل تلك النقرة وربطها بمختلف البيانات المتاحة. بالتالي يمكن اشتقاق بيانات أخرى مثل الإعجاب بكل ما له علاقة بالتكنولوجيا أو بالطبيعة, أو حتى الميولات الأدبية والسياسية. لذلك ، سيكون من الضروري التمييز بين البيانات الحقيقية والبيانات المشتقة ومدى الحاجة لتخزينها ومعالجتها.

هناك دراسة أجراها باحثون من جامعة كامبريدج وجامعة ستامفورد ونشرت في عام 2015. أجريت هذه الدراسة على 86000 متطوع وعلى الإعجابات التي استخدموها. جعل من الممكن معالجة 5 سمات شخصية ، مثل: أن تكون لطيفًا ، أو مندفعًا ، أو الجانب المنفتح لكل مستخدم.

على المستوى الكمي ، تعتمد خوارزميات Facebook على عدد الإعجابات:

- مع 10 لايكات ، تعرف هذه الخوارزميات أنك أفضل من زميل في المكتب.
- مع 70 لايك ، يعرفونك كصديق
- مع 150 لايك ، يعرفونك كأحد أفراد العائلة.
- مع 300 لايك ، يعرفونك كزوج.

في المتوسط ، كل مستخدم لديه "أعجبني" أكثر من 227 مرة.

دعنا الآن ننتقل إلى جانب آخر ، وهو تحديد الموقع الجغرافي. ففي عام 2013 ، أظهرت دراسة أجراها 4 باحثين من مخبر البحوث UMIT بالجامعة السويدية UMEA أن بيانات تحديد الموقع الجغرافي لدينا متكررة ومنتظمة وفريدة من نوعها. من خلال 4 نقاط مكانية وزمانية ، يمكننا تحديد 95٪ من السكان. مثلا،

النقطة الأولى: أين ننام؟

النقطة الثانية: أين نقضي اليوم؟

النقطة الثالثة: أين نعمل؟

النقطة الرابعة: أمام أي مدرسة توقفنا؟

النقطة الخامسة: في أي قاعة رياضية نمارس التمارين الرياضية؟

يمكن أن تكون الأسئلة الأخرى موضوع تجميع البيانات التي تم الحصول عليها من أجل اتخاذ تدابير و / أو قرارات استراتيجية. مثلا: معرفة كيف يتحرك الناس؟ كيف يزور السائحون بلادنا؟ تسهل هذه البيانات ، عندما يتم تجميعها جميعًا ، العمل على تحسين البنية التحتية التقنية واتخاذ قرارات مثل توقع ذروة الزوار.

قبل 15 عامًا ، لم يكن لدى أحد جهاز GPS. اليوم ، مع تداول الهواتف الذكية ، من الممكن الحصول على عدة معلومات عن كل واحد منا. أيضًا ، إذا قمنا بتجميعها ، عن سكان مدينة أو السياح الذين يزورون بلدًا ما تصبح البيانات ذات قيمة مضافة عالية.

نقوم بإنشاء المزيد والمزيد من البيانات غير المتجانسة: الصور والأفلام والتعليقات على الشبكات الاجتماعية، إلخ. نستخدم أيضًا المزيد والمزيد من العناصر المتصلة: الساعات، والأساور، وأعمدة الإنارة في المدن الذكية، والسيارات، وما إلى ذلك. كل هذا الحشد من البيانات غير متجانس. البيانات الضخمة قادرة على احتواء كل هذا التنوع مما يتيح انا إمكانيات هائلة لاستغلالها من أجل إنتاج المعرفة.

7. 4البيانات الضخمة والصدق

يشير الصدق (Veracity) إلى قضية موثوقية البيانات. مع زيادة الكمية ، يتم فقد الجودة والدقة (الاختصارات، الأخطاء المطبعية ، التشوهات ، المصدر غير الموثوق به ، إلخ). يجب أن تعالج حلول البيانات الضخمة هذا الإشكال. من هنا تأتي الحاجة إلى صرامة كبيرة (جدًا) في تنظيم التجميع والتدقيق المتقاطع ، والعبور ، وإثراء البيانات لإزالة عدم يشير اليقين ، وما لا يمكن التنبؤ به. طبيعة البيانات المقدمة في النماذج ولكن أيضًا لاحترام الإطار القانوني لخلق الثقة وضمان أمن وسلامة البيانات.

وفقًا لتقرير[16] يتناول حماية البيانات الخاصة نُشر في عام 2015 من قبل شركة الأمن السيبراني Symantec ، أشار 57٪ من الأوروبيين إنهم قلقون بشأن أمان معلوماتهم الشخصية ، ويعتقد 81٪ أن بياناتهم تساوي أكثر من ألف يورو و 31٪ أصبحو يترددون في توصيل بياناتهم الشخصية خاصة أمام ظهور العديد من التطبيقات المزيفة.

تم تطوير تلك التطبيقات لإنشاء بيانات مزيفة في محاولة لخداع تطبيقات Android التي يشعر المستخدم أحيانًا أنها تطفلية للغاية. Xprivacy هي أداة تغذي تطبيقات Android بجهات اتصال مزيفة ، و إحداثيات جغرافية مزيفة ، وقواميس مستخدم مزيفة ، وحافظات ذاكرة مزيفة ، وسجلات مكالمات مزيفة ، ورسائل SMS مزيفة. الهدف المعلن لـ Xprivacy هو إنشاء بيانات خاطئة لحماية خصوصيتك بشكل أفضل.

16 Teboul, B., Berthier, T., "Valeur et Véracité de la donnée :Enjeux pour l'entreprise et défis pour le Data Scientist", Actes du colloque "La donnée n'est pas donnée", École Militaire, France,2015.

في نفس السياق ، يتيح موقع FakeNameGenerator إنشاء قاعدة بيانات بتنسيقات مختلفة (MS SQL و MySQL و IBM DB2 و Oracle وما إلى ذلك) تضم 50000 هوية وهمية متماسكة بما في ذلك الاسم والعمر والجنسية والعنوان ومهنة الملفات الشخصية المحفوظة. أما بالنسبة لموقع CloneZone ، فهو يقوم بالاستنساخ الفوري لأي موقع ويب ويعرض تعديل أو تحويل كل محتوياته أو جزء منها.

مثال آخر يمكن الاستشهاد به هو ذلك المتعلق بالأمن السيبراني. يتزايد إنتاج مجموعات بيانات وهمية متسقة أثناء مرحلة الهندسة الاجتماعية للتحضير لهجوم إلكتروني. العامل البشري هو بالفعل أول ثغرة يتم استغلالها أثناء هجوم رقمي ، قبل هشاشة الأنظمة. التحدي الرئيسي للمهاجم هو خلق مساحة ثقة بينه وبين ضحيته ومن ثم الاستفادة من هذه الثقة لتنفيذ الهجوم. تزداد "الطاقة" التي يتم نشرها لبناء الثقة بشكل مطرد مع وعي المستخدمين بالمخاطر والمخاطر الرقمية.

الفخاخ الإلكترونية التي تعمل اليوم معقدة بشكل متزايد وتعتمد في بعض الأحيان على هياكل بيانات وهمية متقنة. وهكذا، أظهرت عملية التجسس السيبراني الأخيرة (Cyberespionnage Newscaster) قوة البيانات الخاطئة في خداع أهدافها. ومن بين ضحايا هذا العدوان ضباط كبار في الجيش الأمريكي ومهندسون من صناعة الأسلحة وأعضاء في الكونجرس وقادة أعمال.

7. 5 البيانات الضخمة والقيمة

يجب علينا تحويل جميع البيانات إلى قيمة (Value) قابلة للاستخدام. البيانات التي لا قيمة لها لا فائدة منها. يجب تحقيق الأهداف الإستراتيجية لخلق القيمة لأصحاب المصلحة والمؤسسات في جميع مجالات النشاط.

يعتبر نهج البيانات الضخمة منطقيًا فقط لتحقيق الأهداف الإستراتيجية لخلق القيمة للمؤسسة؛ في جميع مجالات النشاط: التجارة ، الصناعة ، الخدمات ، إلخ. لا يهم نجاح مشروع البيانات الضخمة المستخدمين إلا إذا جلب قيمة مضافة ومعرفة جديدة.

تغطي القيمة عدة أطياف ، يتطلب كل منها تحليلاً محددًا. وبالتالي فإننا نتحدث عن قيمة التأثير على السياق أو قيمة النمذجة أو قيمة التنبؤ أو القيمة الإدارية أو القيمة الاقتصادية أو قيمة إعادة البيع. دعونا الآن نقدم حالتين: الحالة الأولى تخص شركة التوربينات فيستاس VESTAS, إذ مكن تنفيذ تحليل البيانات الضخمة شركة تطوير توربينات الرياح فيستاس VESTAS من تحسين عمليتها في تحديد أفضل المواقع لتثبيت توربينات الرياح الخاصة بها. أدت معالجة البيانات الضخمة إلى زيادة أداء إنتاج الكهرباء وخفض تكاليف الطاقة المرتبطة بها. باستخدام بيانات الأرصاد الجوية الكبيرة ، تستطيع فيستاس وصف سلوك الريح فوق المنطقة المختارة وتقديم حالة عمل دقيقة لعملائها. أدى نظام البيانات الضخمة VESTAS-IBM إلى تقليل وقت الاستجابة بنسبة 97٪ لتوقعات الرياح من عدة أسابيع إلى بضع ساعات فقط اليوم. تم تخفيض تكلفة الإنتاج لكل كيلوواط / ساعة للحرفاء بالإضافة إلى التكلفة وبصمة تكنولوجيا المعلومات المرتبطة بتخفيض أكثر من 40٪ في استهلاك الطاقة.

الحالة الثانية تدور حول حديقة حيوان سينسيناتي (Cincinnati Zoo). ففي مواجهة صعوبات الربحية ، قررت حديقة الحيوانات الأمريكية في سينسيناتي (أوهايو) إجراء معالجة البيانات الضخمة لبيانات الحرفاء وبيانات من أجهزة الاستشعار المنتشرة داخل مناطق الجذب والمباني في المنتزه. أدت عملية تحليل البيانات الضخمة ومعالجتها إلى زيادة إنفاق الزائرين بنسبة 25٪ ، مما أدى إلى زيادة الإيرادات الإضافية بما يزيد عن 350 ألف دولار سنويًا. تم تطبيق الفهم المتعمق لبيانات الحرفاء لتحسين إدارة الموارد البشرية لحديقة الحيوان وأتاح الوقت للموظفين المتاحين الآن لشغل مناصب مربحة أخرى. وهكذا استعادت ميزانية الشركة توازنها.

8. مصادر البيانات الضخمة

يوجد العديد من مصادر البيانات الضخمة. من هذه المصادر تلك الناشئة عن إدارة أحد البرامج سواء أكان برنامج حكومي أو غير حكومي، كالسجلات الطبية الإلكترونية وزيارات المستشفيات وسجلات التأمين والسجلات المصرفية وبنوك الطعام وغيرها. كما نجد المصادر التجارية أو ذات الصلة بالمعاملات والتي تعد مصدرا آخر كالبيانات الناشئة عن معاملات بين كيانين، على سبيل المثال معاملات البطاقات الائتمانية والمعاملات التي تجرى عن طريق الإنترنت بوسائل منها الأجهزة المحمولة. كما أن هناك مصادر معتمدة على شبكات أجهزة الاستشعار وأجهزة التتبع فعلى سبيل المثال، التصوير بالأقمار الصناعية، وأجهزة استشعار الطرق، وأجهزة استشعار المناخ وتتبع البيانات المستمدة من الهواتف المحمولة والنظام العالمي لتحديد المواقع وغيرها يمكن أن تكون أحد مصادر البيانات الضخمة.

هناك نوع آخر من المصادر وهو المتعلق بسلوك المستخدم مثل مرات البحث على الإنترنت عن منتج أو خدمة ما أو أي نوع آخر من المعلومات، ومرات مشاهدة إحدى الصفحات على الإنترنت. وأخيرا، مصادر البيانات المتعلقة بالآراء مثل التعليقات على وسائط التواصل الإجتماعي مثل فيسبوك وتويتر وغيرها.

9. تقنيات البيانات الضخمة

يوجد العديد من الأدوات والتقنيات التي تستخدم لتحليل البيانات الكبيرة مثل HPCC, Hadoop Map Reduce, إلا أن (Hadoop)يُعد من أشهر هذه الأدوات. Hadoop هو برنامج أو منصة برمجية مفتوحة المصدر مكتوبة بلغة البرمجة جافا (Java), تستخدم لتخزين ومُعالجة البيانات الضخمة بشكل موزع. أي أن تخزين هذه البيانات الضخمة تكون على عدة أجهزة ومن ثم توزع عملية المعالجة على هذه الأجهزة

لتسريع العملية. ثم تعود أو تستدعى كحزمة واحدة. تتكون الأدوات التي تتعامل مع البيانات الضخمة من ثلاثة أجزاء رئيسية وهي: أدوات التنقيب عن البيانات(Data Mining) وأدوات التحليل(Data Analysis) وأخيراً أدوات عرض النتائج(Dashboard).

10. الأطراف فى منظومة البيانات الضخمة

لكي يتم تنظيم أي خدمة يجب تحديد الأطراف التي تتعامل مع هذه الخدمة وتحديد واجبات وحقوق كل طرف. تتكون منظومة البيانات الضخمة من عدة جهات تتفاعل فيما بينها ،هذه المنظومة مُكونة من موفر البيانات الضخمة (Big Data Provider) ومُقدم خدمة البيانات الضخمة (Big data service provider)وعميل خدمة البيانات الضخمة (Big data service customer). في ما يلي تفصيل لكل طرف متدخل:

أولا: موفر خدمة البيانات الضخمة

يعمل موفر البيانات الضخمة على توفير البيانات من مصادر مُختلفة إلى مُقدم الخدمة، وتشمل أنشطة موفري البيانات على سبيل المثال إنشاء البيانات وإنشاء المعلومات الوصفية (Meta data) التى تصف مصدر البيانات وإيجاد مصادر البيانات المفتوحة (Open data) على الإنترنت وتوفير بيان الخدمات (Service catalogue) إلى مُقدم الخدمة عن البيانات القابلة للإستخدام.

ثانياً: مُقدم خدمة البيانات الضخمة

يقوم مُقدم الخدمة بتحليل البيانات الضخمة وتوفير البنية التحتية اللازمة لها. وتشمل أنشطة مُقدم الخدمة على سبيل المثال البحث فى مصادر البيانات, وجمع البيانات عن طريق الطلب المباشر من موفر البيانات أو البحث فى الإنترنت (DataCrawling). كما تشمل أنشطة مقدم الخدمة تخزين البيانات ودمجها وتوفير الأدوات لتحليلها ودعم إداراتها من خلال أدوات التصرف في خصوصية البيانات وأمن البيانات وملكية البيانات وغيرها.

ثالثاً: حريف خدمة البيانات الضخمة

هو المستخدم النهائي لمنظومة البيانات الضخمة. وهو نظام يستخدم النتائج أو الخدمات التى يُقدمها مُقدم خدمة البيانات الضخمة. كما يمكن للعميل أن يُنتج خدمات جديدة أو معرفة وذلك إعتماداً على نتائج تحليل البيانات الضخمة. تشمل أنشطة العميل على سبيل المثال طلب خدمة البيانات الضخمة من مزود الخدمة واستخدام مُخرجات خدمة البيانات الضخمة في النشاط الذي يرغبه ويتخصص فيه.

11. تطبيقات البيانات الضخمة

11. 1 القطاع الحكومى والبيانات الضخمة

عندما يتعلق الأمر بإدارة البيانات، أغلب المنظمات الحكومية تواجه مشكلة وجود كميات هائلة من البيانات في أنظمة الكمبيوتر. ومعظم هذه البيانات غير منظمة أو مُهيكلة (Unstructured Data) وهذا يعني أنها لا تناسب أي نموذج بيانات معرّف مسبقاً. لفهم الأنماط الموجودة في هذه البيانات يجب أن تطبق المنظمات الحكومية نماذج إحصائية تسعى لإلتقاط ومعالجة كميات هائلة من البيانات غير المهيكلة وتسمى هذه العملية

بالبيانات الضخمة. نشير أيضا إلى أن أغلب المنظمات الحكومية لا تملك عدد كافي من الموظفين أو لا تملك القدرة الحسابية اللازمة.

لإدارة وتحليل جميع البيانات الخاصة بهم، ومع طبيعة البيانات المتغيرة وارتفاع حجمها أصبحت الإستعانة بأدوات البيانات الضخمة من خلال الحوسبة السحابية (Cloud Computing) أمراً ضرورياً. فأصبح بإمكان المختصين بتطوير الخدمات الحكومية مثلا رصد مدى رضاء المواطنين عن الخدمات المقدَّمة لهم. وعلى ضوء النتائج المحللة يمكن استنتاج ما يلزم عمله للتطوير والتحسين. حيث أصبح مسح آراء الجمهور عن طريق الاستبيانات التقليدية مكلفاً وغير مجدٍ في كثير من الأحيان، وذلك نظراً لتنوع البيانات الديموغرافية وثقافات المتعاملين. إن من أكبر المصادر لتلك البيانات الضخمة هي البيانات المسجلة من خلال عمليات التعداد السكاني والتسجيل في قواعد البيانات الحكومية، حيث يمكن أن تستنتج الحكومات معلومات ثمينة جداً من خلال تحليل تلك البيانات المخزنة.

11. 2 تحليل البيانات الضخمة و تحسين التعليم

أدى إستخدام أدوات التعلم عبر الإنترنت والبرامج القائمة على التفاعل بصورة متزايدة في مجال التعليم إلى زيادة حجم البيانات، واختلاف نوعية البيانات الكبيرة التي يُمكن جمعها من بيئات التعلم. إذ نجد بيانات كبيرة عن المتعلمين وخبرات التعلم لدى المتعلمين. كما نجد بيانات متعمقة داخل بيئات التعلم، وبيانات حول التفاعلات الإجتماعية في بيئات التعلم، وبيانات مُفصلة عن أنشطة التعلم من نصوص ووسائط ومقاطع فيديو وغيرها. كما تختلف هذه البيانات في نوعيتها وعمقها بنسب متفاوتة.

يمكن الاستفادة من تحليل هذه الأنواع من البيانات الضخمة في التعليم، لتوفير مجموعة مُتنوعة من الفرص والخيارات بهدف تحسين تعلم الطلاب من خلال التعلم التكيفي أو التعليم القائم على الكفاءة. هذا مما ينتج عنه

تعلم أفضل نتيجة لتشخيص أسرع وأكثر تعمقاً في بيانات حقيقية تراكمية لاحتياجات المتعلم أو المتاعب التي تواجهه أثناء عملية التعلم. ويندرج ضمن ذلك تقييم المهارات مثل التفكير المنظم، والتعاون، وحل المشاكل في سياق عميق، وتقييم أصيل لمجال وموضوع المعرفة، بالإضافة لتحديد التدخلات المستهدفة لتحسين نجاح الطلاب وخفض التكاليف الإجمالية للطلاب والمؤسسات، واستخدام البيئات القائمة والمعلومات المعقدة في صنع القرارات وتحديد السياسات. ويمكن أن توفر هذه البيانات أدوات حديثة وفعالة لقياس أداء الطلاب للمهام التعليمية. ويمكن أن تُساعد كذلك في تصميم بيئات تعلم تصميماً مُخصصاً وفق إحتياجات مُحددة للطلاب. كما يمكن أن تعطي تحليلاً واضحاً لردود الفعل الفردية والجماعية لمجموعة من القضايا التعليمية وغيرها من المميزات.

11. 3 تحليل البيانات الضخمة لتحسين عملية صنع القرار

إن عملية اتخاذ القرارات تُعد محور العملية الإدارية وجوهرها وإن نجاح المؤسسة أو القطاع الحكومي يتوقف إلى حد كبير على قدرة وكفاءة القيادة الإدارية على إتخاذ القرارات الإدارية المناسبة. إن عملية صنع القرار تبدأ بتجميع البيانات ومُعالجتها واستخلاص المعلومات التي بناء عليها يتم اتخاذ القرار حيث بدأت تعتمد العديد من الشركات الكبيرة والقطاعات الحكومية على سياسة تحليل البيانات الضخمة والمعقدة والتي تحتاج إلى البرمجيات المتخصصة في مجال إدارة البيانات والتحليلات، والتي لا يمكن مُعالجتها باستخدام أداة واحدة فقط أو العمل على تطبيقات مُعالجة بيانات تقليدية، فمن المعروف أن جمع البيانات والمعلومات تُساعد على التوصيف الدقيق للمشكلة وتحليلها للوصول إلى نتائج دقيقة، لذلك كان لابد من إعتماد نظام إداري يشمل تحليل البيانات الضخمة والهائلة جداً. يستخدم القطاع الحكومي والشركات الكبيرة نظام تحليل البيانات الضخمة لتحسين العمليات الداخلية، مثل إدارة المخاطر، إدارة علاقات العملاء، والخدمات اللوجستية. كما يستخدم لتحسين المنتجات والخدمات

القائمة، وتطوير الخدمات والمنتجات الجديدة والإستفادة من المعلومات وتقديم العروض المناسبة للمستفيدين في الوقت المناسب.

11. 4 الاستفادة من البيانات الضخمة في المجال العسكري

البيانات الضخمة تتيح فرصاً عديدة للصناعة العسكرية، وخصوصاً أن تحليل البيانات الضخمة يسمح بالكشف عن دلالات تنفيذية يرتكز عليها صنّاع القرار لتطوير مختلف الشؤون العسكرية. إذ تساعد البيانات الضخمة مثلا على تطوير قدرات الاستخبارات العسكرية من خلال جمع البيانات من مصادر مختلفة وبناء منصة حاسوبية مترابطة تعزّز تبادل المعلومات بين العسكريين.

يُشكّل النظام الذي يتمتع بشبكة إتصال كبيرة بين الأشياء المتصلة (Connected things) أداة مهمة لتوليد شبكة المعلومات في المجال العسكري. فيساعد على جمع وتبادل المعلومات بوتيرة أسرع و على تحليلها بكفاءة وفعالية. لقد عزز إنترنت الأشياء (IoT) النظام العسكري في ساحات المعارك من خلال تبادل المعلومات، وتحديد مواقع العدو على أرض المعركة وغيرها. بحيث بات يُشكل تحليل المعلومات عن العدو بشكل دقيق وبالوقت الفعلي إحدى الخطوات الأساسية في عملية صنع القرار العسكري. إذ تظهر تكنولوجيا البيانات الضخمة قدرة على تحليل هذه المعطيات بشكل صحيح، مايسهل عملية اتخاذ القرار من قبل القادة العسكريين.

11. 5 الاستفادة من البيانات الضخمة في المجال الاقتصادي

لقد أصبح بإمكان الشركات والمؤسسات والهيئات اليوم على إختلاف أنواعها تحليل حركة الحرفاء من شراء وبيع ونحوه بدقة أكبر ليتمكنوا وفقاً لذلك من معرفة السلع الأكثر طلباً أو تلك الراكدة. ثم بناءا عليه يقترحون على

عملائهم سلع معينة وفقاً لعمليات الشراء التي تتم. كما أصبح لديهم القدرة على فهم سلوك المستفيدين بشكل أكثر دقة وتحديد المميزين منهم ومن هم بحاجة للمساعدة أو تحديد توجهاتهم أو مُراقبة أدائهم.

هذا الأمر ليس فقط لمراكز البيع التقليدية بل يشمل المتاجر الإلكترونية على شبكة الإنترنت وعلى نطاق أوسع. فأصبح يتفاجأ مُستخدم شبكات التواصل الإجتماعي أو البريد الإلكتروني في أحيان كثيرة، بظهور إعلانات تجارية لسلع قام مسبقاً بالبحث عنها في تطبيقات أخرى. بل أكثر من ذلك هناك بعض الخوارزميات التي تستخدم بيانات تحديد الموقع على جهاز الهاتف لاقتراح الإعلانات. ومن هنا نجد أن ذلك يحدث نتيجة لتحليل البيانات الضخمة الناتجة من هذه المواقع والإستفادة منها في التسويق وذلك باستخدام كل جزء صغير من البيانات المتاحة عن المستخدمين لمعرفة ميولهم وتفضيلاتهم بغية عرض البضائع بأمثل طريقة ممكنة تجلب لشركات التسوق الإلكتروني أعظم ربح ممكن. ولا تقوم بتحليل البيانات التي تحصل عليها من تصفح المستخدم للإنترنت وحسب، بل قد تتبع زيارته للأسواق الحقيقية من خلال جهاز تحديد المواقع المثبت بجهازه.

ربما يعتقد البعض أن هذا إختراق للخصوصية، لكن في حقيقة الأمر فإن الشركات التي تقوم بذلك قد حصنت نفسها من خلال اتفاقية الشروط التي يوقع عليها المستخدم عند تسجيل الدخول إلى تطبيقات التواصل الإجتماعي كالفيسبوك وتويتر.

قد دخلت عملية التجارة الإلكترونية في السنوات الأخيرة مستوى جديداً من التنافس، بظل الكم الهائل من بيانات المستخدمين التي توفرها شبكات التواصل الاجتماعي واستخدام الإنترنت. أصبح الشاغل الأول لعمالقة التجارة الإلكترونية هو كيفية البقاء في المنافسة. فمن ناحية، تحتاج الشركات للترويج للبضائع وهذا يتطلب معرفة باحتياجات الزبائن، ومن ناحية أخرى تحتاج لأن تقدم بضائعها بأسعار تنافسية مع المحافظة على هامش ربح كبير.

11. 6 الاستفادة من البيانات الضخمة في المجال الطبي

تحول البشر شيئاً فشيئاً من طب التعامل مع الأمراض إلى طب يسعى للتنبؤ بالأمراض ومنعها. فكما يقال الوقاية خير من العلاج. هذا ما يفتح الباب لعصر جديد من الطب تلعب فيها البيانات الصحية الضخمة وتحليلاتها دوراً بارزاً.

أصبح بإمكان المستشفيات الحكومية والمراكز الطبية والأطباء الإستفادة من البيانات الضخمة في دراسة سلوكيات المرضى عبر تحليل ملفاتهم الطبية والزيارات التي قاموا بها للعلاج والتقنيات القابلة للارتداء مما قد يساعدهم على تقديم خدمات طبية أفضل. يتيح تحليل البيانات الضخمة معالجة أوجه القصور في نظم تقديم الرعاية الصحية التي تتزايد تكاليفها بفعل النمو السكاني وارتفاع متوسط الأعمار. كما تستفيد العديد من المستشفيات في مختلف أنحاء العالم من البيانات الضخمة في تقليل وقت الانتظار في أقسام الطوارئ وتتبع حركة المريض، وزيادة كفاءة الإدارة الطبية.

كما تستخدم البيانات الضخمة في صناعة الأدوية وتوزيعها وبيعها. حيث تجمع شركات تصنيع الأدوية والتأمين الصحي البيانات من عدة مناطق جغرافية لاستخدامها في التنبؤ بظهور أمراض معينة، وزيادة مبيعاتها في مناطق معينة. كما تعتمد سياسات التسعير وتوزيع الأدوية على نتائج تحليل هذه البيانات.

12. الخاتمة

أصبحت تكنولوجيا المعلومات توظف أدواتها لتحليل ومعالجة البيانات الضخمة (Big Data) والتي تهدف الى استثمارها في العديد من المجالات ومنها المجال العسكري والمجال الطبي والقطاع الحكومي والاقتصادي والتعليمي

وغيرها .حيث تتم عملية استرجاع البيانات ومعالجتها واستخدامها بغرض تطوير المنتجات أو استحداث منتجات جديد، أو توفير المعلومات اللازمة لمساعدة متخذي القرار لصناعة مخرج مناسب وتطوير عملية التسويق الإلكترونى وتطوير الأدوية والمساعدة فى إكتشاف الأمراض وتحقيق الأمن القومي وتحسين العملية التعليمية.

تحليلات البيانات

1. مقدمة

منذ سنوات تعدد الحديث عن مفهوم الذكاء في النظم التي تعتمد على المعلوماتية مثل المدن الذكية (Smart cities) و الرياضة الذكية (Smart sport) والمنازل الذكية (Smart home). وتعتبر البيانات الضخمة في هذا السياق القلب النابض لبناء مثل هذه الأنظمة الذكية. حيث أن هذه النظم تعتمد اليوم أساسا على جمع البيانات وتخزينها وتحليلها من أجل استخراج المعارف التي تسمح بإنتاج مؤشرات تسمح بالاستغلال الأمثل لهذه الأنظمة.

ولقد سمح التقدم التكنولوجي بحدوث نقلة نوعية في مجال تحليل البيانات خاصة مع انتشار الحوسبة السحابية (Cloud Computing) و خوارزميات تعلم الآلة (Machine Learning) وتكنولوجيات ذات العلاقة بالتصرف ومعالجة البيانات الغير مهيكلة و النصف مهيكلة (Unstructured and Semi-structured data).

ويهدف علم تحليل البيانات إلى معالجة وتحليل واستخراج المعرفة, من خلال استثمار البيانات بمختلف أشكالها وصيغها ومصادرها, من أجل التنبؤ أو رصد أو اقتراح ما من شأنه أن يقدم خدمة ذات قيمة مضافة للمستخدم. هذه القيمة المضافة لا يمكن إدراكها مباشرة من خلال الاطلاع على البيانات الخام بسبب حجمها وتداخلها وتنوعها.

2. استراتيجية بناء منظومة تحليل البيانات

لبناء منظومة تحليل بيانات يعمد الخبراء أولا إلى تحديد الهدف من بناء هذه المنظومة. أي تحديد الغاية المنشودة ونوعية النتائج المتوقعة منها. انطلاقا من هذه النقطة يتم دراسة وتحديد نوعية المعطيات المتوفرة وذات الصلة التي تمثل المادة الخام التي تحتوي في ثناياها على المعارف المنشودة. المرحلة التالية هي عملية تحليل البيانات التي تعتمد على جملة من التقنيات منها استخدام خوارزميات تعلم الآلة. الهدف منها هو تحويل الكم الهائل من المعطيات إلى معارف يمكن استعمالها واستغلالها من أجل الإجابة على أسئلة أو الحصول على توقعات أو إنتاج مؤشرات تسمح بإدراك أعمق وفهم أشمل للبيانات. هذا يتطلب الاعتناء أيضا بطريقة تقديم النتائج المستخلصة من البيانات إثر عملية التحليل. هذه المرحلة هي مرحلة إعداد التقارير التي ستقدم للمستخدم النهائي.

عادة يتم التعويل عادة على خبراء تحليل البيانات (Data analyst) لبناء استراتيجية بناء منظومة تحليل البيانات. أما التنفيذ فيتعهد به خبراء البرمجة, حيث يتم الإستفادة من مختلف التكنولوجيات المتوفرة من أجل بناء منظومة تحليل البيانات.

3. مصادر البيانات

لعل أول سؤال يتبادر إلى الذهن عند بناء منظومة تحليل بيانات يتعلق بمصدر هذه البيانات. هنا, أول مسألة وجب التنويه إليها هي أن البيانات ليست ضرورة بيانات داخلية بل يمكن أن تكون بيانات خارجية. فما هو المقصود بالبيانات الداخلية والخارجية؟

فيما يتعلق بالبيانات الداخلية فالمقصود بها تلك البيانات المتوفرة في المنظومة المعلوماتية بصفة ذاتية والتي تتمحور حولها مختلف البرمجيات. فعلى سبيل المثال, إذا تعلق الأمر بمنظومة معلوماتية تهتم بإدارة الموارد البشرية, فالبيانات الداخلية هي تلك التي تهم الموظفين من بيانات شخصية, وبيانات مهنية وغيرها من البيانات ذات الصلة. في نفس السياق فالنفترض أن مسؤولي إدارة الموارد البشرية يرغبون في استشراف العطل المرضية. البيانات الداخلية تحتوي أرشيف العطل المرضية للموظفين, وهي بيانات مهمة تسمح بالقيام بعملية التحليل التي تسمح بالاستشراف. لكن ربما يمكن تحسين دقة الاستشراف من خلال إضافة بيانات الطقس, هذه البيانات ليست من ضمن البيانات الداخلية ولكنها بيانات يمكن الحصول عليها من أطراف خارجية تقدم هذه الخدمة.

إذا البيانات الخارجية هي بيانات مفيدة وربما أساسية لعملية تحليل البيانات ولكن مصدرها خارجي. قد تكون متوفرة من خلال خدمات مجانية أو من خلال خدمات بمقابل. الفكرة الأساسية هنا هو أنه لا يجب الاقتصار على البيانات الداخلية عندما يتعلق الأمر بتحليل البيانات.

بالعودة أيضا إلى السؤال الذي طرحناه حول مصدر البيانات فيجب أيضا أن ننوه إلى أن البيانات تجمع بعدة طرق منها المباشرة والتي تستدعي تدخل المستخدم لرقنها مباشرة, ولكنها قد تأتي بصفة غير مباشرة من خلال جمع آثار الاستخدام أو من خلال بيانات ناتجة عن عملية تحليل لبيانات أخرى, وهكذا دواليك.

فلنأخذ على سبيل المثال منصة YouTube كنموذج لمنصة تستعمل تقنية تحليل البيانات. أولا الهدف من تحليل البيانات في هذه المنصة يهدف إلى الإجابة عن سؤالين:

.كيف يمكن إبقاء المستخدم أطول وقت على المنصة وتشجيعه على استعمالها والرجوع إليها؟

.كيف يمكن رفع المداخيل من خلال اقتراح الإعلانات المناسبة للمستخدم؟

السؤالين متعلقين بمسألة توقع المحتوى الذي من شأنه أن يتماشى مع ميولات واهتمامات مستخدم المنصة, وهذا يتطلب بيانات وتحليل هذه البيانات. الوضعية الأولى التي يمكن دراستها هي حين يستعمل المستخدم الموقع دون أن يكون مسجلا أو دون التعريف بنفسه. في هذه الوضعية يتم اللجوء إلى المعطيات الخارجية والتي تتمثل في عنوان الIP للجهاز المستعمل ومن خلال المعطيات الخاصة بالتطبيق أو متصفح الويب المستعمل. هذه المعطيات الخارجية تسمح بالتعرف إلى الموقع الجغرافي المفترض للمستخدم وإلى اللغة التي يستعملها. بناءا على هذه المعطيات تتم عملية تحليل البيانات لاقتراح الفيديوهات الأكثر مشاهدة والأكثر شعبية بين المستخدمين الذين ينتمون لنفس المنطقة الجغرافية و الذين يستخدمون نفس اللغة. هذا التحليل يسمح أيضا باختيار الإعلانات الأنسب. وبما أن هنالك دائما هامش للخطأ في تحليل البيانات, فواجهة المستخدم لمنصة Youtube تسمح بتغيير الموقع الجغرافي واللغة حسب الطلب.

كما أن البيانات حول قوة الربط بالنت ودقة شاشة المستخدم وبيانات المستشعرات, كاتجاه مسك الهاتف الذكي, وحتى ملفات تعريف الارتباط التي يحويها المتصفح, كلها بيانات خارجية ليست محفوظة من قبل الموقع ولكنها متاحة ويتم استغلالها أحسن استغلال.

أما الوضعية الثانية فهي حين يقوم المستخدم بالتعريف بنفسه. حينها يمكن استخدام المعطيات الداخلية التي يحتفظ بها الموقع عنه. وهذه المعطيات تنقسم بدورها إلى ثلاثة أصناف:

. معطيات أدخلها المستخدم, كالاشتراك في قنوات بعينها

. معطيات مأخوذة من آثار استخدام الموقع, كالنقر على فيديوهات بعينها

. معطيات أنتجتها عملية تحليل البيانات السابقة, كتشابه البيانات مع بيانات مستخدمين آخرين

4. أصناف البيانات

لا يمكن الحديث عن تحليل البيانات دون التعريج على أصناف البيانات وتبيان ما يميزها عن بعضها. فالبيانات التي يتم جمعها يمكن تقسيمها إلى ثلاثة أصناف:

. البيانات المهيكلة (Structured data)

. البيانات الغير مهيكلة (Unstructured data)

. البيانات النصف مهيكلة (Semi-structured data)

أما البيانات المهيكلة فيمكن تعريفها بأنها بيانات ذات بنية ثابتة. هذه البيانات هي التي نجدها عادة في جداول البيانات (Spreadsheets) وقواعد البيانات العلائقية (Relational databases) حيث تتكون بنيتها من حقول ثابتة. كل حقل بيانات يحمل إسما يبين دلالته. إضافة إلى ذلك يمكن, حسب التكنولوجيا المعتمدة لحفظ والتصرف في البيانات, تحديد نوع البيانات التي يمكن أن يحتويها كل حقل بيانات ومختلف القيود المتعلقة به.

فعلى سبيل المثال لو سلمنا بأن البيانات التي نود حفظها ومعالجتها تهم منتج ما, فهذه البيانات هي بيانات مهيكلة بما أن بيانات كل المنتجات هي ذات بنية ثابتة تحوي عددا ثابتا من الحقول. على سبيل المثال يمكن أن تكون هذه الحقول هي معرف المنتج, اسم المنتج, سعر المنتج, وكمية المنتج المتوفرة. ثم يمكن اضافة جملة من القيود على كل حقل. فلو أخذنا على سبيل المثال الحقل الخاص بالسعر, فيمكن تقييده بأن المحتوى يجب أن يكون عددا عشريا, ويمكن حتى تحديد عدد الأرقام بعد الفاصلة, كما يمكن تقييده بأنه عدد أكبر من الصفر بالضرورة. وقس على ذلك بالنسبة لبقية الحقول المكونة للبيانات المهيكلة. البيانات المهيكلة هي الصنف التقليدي من البيانات وتتميز بأنها سهلة التخزين, سهلة المعالجة, وبالتالي فهي مناسبة لعمليات تحليل البيانات.

البيانات الغير مهيكلة هي ,بعكس البيانات المهيكلة, تفتقر لهيكلة ثابتة تسمح بتخزينها في شكل جداول متكونة من حقول ثابتة للبيانات كما هو الحال عند التخزين في جداول البيانات وقواعد البيانات العلائقية. كما أن هذه البيانات قد تأتي مشفرة وفق نظام تشفير ما. من ضمن الأمثلة الأكثر شيوعا للبيانات الغير مهيكلة نجد الصور الرقمية والفيديوهات والصوتيات والملفات الرقمية عموما. هذا النوع من البيانات أصبح اليوم شديد الشيوع, فمنه الصور الطبية, وصور الأشخاص, وفيديوهات المراقبة على سبيل المثال. لاريب أن الحاجة لتحليل مثل هذه البيانات أصبح ضروريا في عدة تطبيقات. في المقابل هذه النوعية من البيانات يصعب تحليلها مما يستلزم عادة جملة من عمليات المعالجة الرقمية لإضفاء بعض الهيكلة عليها أو لاستخراج بعض البيانات منها, كالتعرف الآلي على الكتابة من خلال تحليل الصور وغيرها من التقنيات المماثلة.

فيما يخص البيانات النصف مهيكلة, فهي تفتقر إلى هيكلة ثابتة مع توفر أوسمة (Tags) تسمح بتحديد الحقول المكونة للبيانات. إذا هناك حد أدنى من الهيكلة ولكنها متغيرة مما لا يسمح بتخزينها ومعالجتها من خلال جداول

البيانات أو قواعد البيانات العلائقية كما هو الحال مع البيانات ذات الهيكلة الثابتة. في المقابل فإن هذا الجانب الغير الثابت من هيكلة البيانات يعطيها قدرا من المرونة التي تجعلها واسعة الانتشار.

فعلى سبيل المثال, صفحات الويب هي نصف مهيكلة بما أن كل صفحة لديها هيكلة مختلفة تحددها الأوسمة الموجودة في النص المصدري (Source code). هذه الأوسمة تسمح بالتعرف على مكونات الصفحة من عناوين وفقرات وغيرها من العناصر. وهذه النوعية من البيانات نجدها محفوظة إما في ملفات من نوع XML ومشتقاته (XHTML, RDF, ...) أو ملفات من نوع JSON وهو اختصار ل JavaScript Object Notation, أو في قواعد البيانات من نوع NoSQL. انتشار البيانات نصف المهيكلة يجعلها أيضا من أصناف البيانات التي تستخدم على نطاق واسع في عملية تحليل البيانات.

5. أنواع تحليلات البيانات

الهدف من تحليل البيانات هو الحصول على فهم عميق لما تحمله من معارف, حيث كما أشرنا سابقا هذا الفهم هو أعلى قيمة من البيانات الخام وأكثر إفادة للمستخدم. في هذا الإطار تنقسم تحليلات البيانات إلى عدة أنواع, سنتناول أهمها في ما يلي.

1 .5 تحليل النصوص

يتعلق تحليل النصوص بعملية فهم المحتويات النصية المرقمنة وذلك عادة في إطار واحدة من العمليات التالية :

-تصنيف النصوص (Text categorization)

- تجميع النصوص (Text clustering)

- استخراج المفاهيم (Concept extraction)

- تحليل المشاعر (Sentiment Analysis)

- تلخيص الوثائق (Document Summarization)

فيما يتعلق بعملية تصنيف النصوص (Text Categorization) فهي تهدف إلى تحليل بيانات النص قصد تصنيفه. بالتالي يمكن مثلا في هذا الإطار أتمتة تصنيف النصوص حسب اللغة. كما أن من الاستعمالات الشائعة لهذه التقنية تصنيف الرسائل الالكترونية سليمة أو مزعجة (Spam). هذا على سبيل الذكر لا الحصر لعدد من استخدامات تحليل النصوص من أجل تصنيفها.

أما في خصوص عملية تجميع النصوص (Text clustering) فالمقصود هو تجميعها حسب الموضوع الذي تتناوله. هذا النوع من التحليل الآلي يستخدم مثلا في بعض محركات البحث عندما تكون الكلمة المستخدمة للبحث ذات معاني متعددة. فكلمة على غرار كلمة "خلية" قد توجد في نصوص تتحدث عن مواضيع متعددة تستوجب عملية تجميع. فينتج عن هذه العملية استخراج مجموعات من النصوص حسب الموضوع انطلاقا من نفس الكلمة (خلية): تربية النحل, علم الأحياء, الأمن, ...

استخراج المفاهيم (Concept extraction) هي إحدى عمليات تحليل النصوص وذلك قصد استخراج المفهوم الذي تحمله لفظة ما في نص بعينه. فمعلوم أن الألفاظ حمالة أوجه, بالتالي فإن السياق الذي وردت فيه من ناحية والكلمات المحيطة بها من ناحية أخرى هي المفاتيح لفهم المعنى الذي تحمله. عملية استخراج المفاهيم هي مسألة يسيرة نسبيا بالنسبة للبشر ولكنها غاية في التعقيد والصعوبة بالنسبة للآلة. لذلك تحتاج خوارزميات التحليل الآلي للبيانات النصية المستعملة في استخراج المفاهيم إلى عملية تدريب باستعمال عدد كبير من النصوص الموسَّمة (Tagued) سلفا. الهدف هو توفير عدد كبير من الأمثلة لهذه الخوارزميات من أجل تقليص هامش الخطأ. أما في ما يخص استعمالات استخراج المفاهيم فهي أيضا متعددة, كما هو الحال بالنسبة للآليات الأخرى. فهي تستعمل كثيرا في البحث في الوثائق والنصوص القانونية وذلك لتعدد الكلمات والألفاظ ذات المعاني المتعددة.

تحليل المشاعر (Sentiment Analysis) أو ما يطلق عليه أيضا استخراج الآراء (Opinion mining), هو إحدى عمليات تحليل البيانات النصية والتي تهدف إلى تحديد إذا ما كان النص أو جزء منه أو جملة منه تحمل صيغة إيجابية, أو صيغة سلبية أو صيغة محايدة. كما يمكن للتحليل أن يذهب أبعد من ذلك في استقراء المشاعر التي يحملها النص من قلق أو غضب أو فرحة أو غيرها من المشاعر. ومن جملة الاستعمالات الشائعة لهذه العملية يمكن ذكر تحليل تعليقات المتابعين على منصات التواصل الاجتماعي تفاعلا مع منتج تجاري أو مع موقف لشخصية عامة, وذلك قصد تبين رأي وموقف المتابعين أوتوماتيكيا من خلال تحليل المشاعر التي يحملها التعليق النصي.

تلخيص الوثائق (Document Summarization) يمثل أيضا إحدى آليات تحليل البيانات النصية. هذه الآلية تهدف إلى تحديد أهم المحتويات النصية في وثيقة ما من أجل إنتاج ملخصات أوتوماتيكية. من الاستعمالات المعروفة لهذه الآلية نجد الملخصات الآلية لأهم الأخبار والمستجدات التي تنتجها بعض محركات البحث دون حاجة المستخدم لقراءة الوثيقة الأصلية.

5. 2 تحليل الخطاب

كما هو الحال مع النصوص المكتوبة فإن من جملة البيانات التي يمكن تحليلها اليوم هي البيانات الصوتية. الأهداف من هذا التحليل متنوعة, فقد يهدف التحليل إلى تحديد مشاعر المتكلم, أو إلى تحديد موضوع المحادثة وغيرها من الاستخدامات. نتحدث هنا طبعا عن تحليل الخطاب (.(Speech analytics

لقد شهد التعرف إلى معنى الكلام وتحليله انتشارا واسعا خاصة مع انتشار تطبيقات تلقي الأوامر الصوتية الموجودة في الهواتف الذكية والسيارات وغيرها من الأجهزة الرقمية الحديثة. ولعل المفتاح الأساسي لانتشار هذه التكنولوجيا وامكانية التعلم الآلي التي تسمح بأتمتة هذه التطبيقات هو توفر مادة خام هامة من التسجيلات الصوتية المرقمنة. حيث كانت مسألة التخزين تمثل تحديا حقيقيا أمكن تجاوزه بفضل تقنيات البيانات الضخمة والحوسبة السحابية.

5. 3 تحليل الصور والفيديوهات

كما هو الحال مع تحليل الخطاب, فالصور والفيديوهات شكلت طويلا تحديا حقيقيا نظرا لإشكاليات الحفظ والمعالجة. لذلك تم الاعتماد لفترة طويلة على البيانات الوصفية (Metadata) النصية كوسيلة من أجل فهرسة الصور والفيديوهات قصد تمكين المستخدمين من البحث عنها وتصنيفها. وتتضمن البيانات الوصفية النصية معطيات كالكلمات المفاتيح والنصوص التي تصف المحتوى وغيرها من البيانات التي ترتبط أساسا من حيث الكم والنوع بنوعية الاستخدام المستهدف.

لكن مع التطور التكنولوجي وانخفاض تكلفة تخزين ومعالجة البيانات كالصور والفيديوهات من ناحية, ومن ناحية أخرى مع انتشار خوارزميات الذكاء الاصطناعي القادرة على تحليل محتوى هذه الصور والفيديوهات وليس فقط بياناتها الوصفية, أمكن الحديث عن تحليل حقيقي لهذه النوعية من البيانات (Video and Image analytics).

إضافة لهذا المحفز الأول لتطور هذه التكنولوجيا, فإن المحفز الثاني هو تعدد مجالات استعمالها والحاجة الملحة لها. فمن ضمن التطبيقات الأكثر انتشارا لتحليل الصور والفيديوهات يمكن ذكر قضية التعرف على الأشكال ومنها التعرف على الوجوه (Face recognition). أيضا يتم تطبيق هذه التكنولوجيا في مجال تحليل السلوكيات. حيث تعمد بعض المتاجر إلى التحليل الأوتوماتيكي لسلوكيات الحرفاء داخل المتجر اعتمادا على فيديوهات المراقبة. أما التطبيق الآخر الأكثر شيوعا فهو ذلك الذي يهدف إلى الوعي ببعض الوضعيات والتنبيه إليها من خلال تحليل الصور والفيديوهات, ومنها بالخصوص الصور الطبية مما يسمح بالتشخيص المبكر لبعض الأمراض والعلل.

6. عرض نتائج تحليل البيانات

تحليل البيانات لا تكون مفيدة إلا إذا أمكن استثمار نتائجها, وهذا مرتبط بطريقة عرضها التي يجب أن تكون مختصرة وملخصة وواضحة ومفهومة. بالتالي فإن مسألة عرض البيانات هي مسألة أساسية. ولعل الوسيلة التقليدية لعرض البيانات وأكثرها رواجا هي التقارير التأليفية ولوحات القيادة والتي تطورت بشكل كبير.

في الأثناء وجب التنويه إلى أن تعدد التقارير وطولها والخوض في التفاصيل التقنية وعدم الاهتمام بالشكل إضافة للمحتوى قد يكون له نتائج سلبية. بل قد تكون النتائج حتى كارثية في بعض الأحيان كما وقع مع وكالة الفضاء الأمريكية NASA مع عملية إطلاق المركبة الفضائية Challenger سنة 1986. إذ أن فريق من المهندسين تفطنوا إلى مواطن خلل تتعلق بالمركبة من شأنها تهديد سلامتها فأصدروا تقريرا في الغرض. غير أن طول التقرير وتعدد التفاصيل وعدم وضوح النتائج المستنتجة لم يسمح للمشرفين التفطن لهذا الخطر المحدق بالمهمة. وكانت النتيجة انفجار المركبة بعد ثوان من انطلاقها مما أدى لوفاة كل طاقمها المكون من سبعة رواد فضاء.

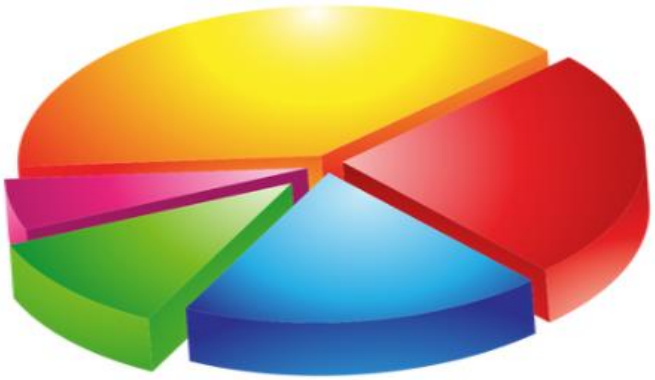
إذا من المهم أن أي نتيجة يتم التوصل إليها من خلال تحليل البيانات يتم توثيقها بطريقة تسمح للأشخاص المعنيين الحصول على المعلومة المناسبة وفق الشكل المناسب حتى يتمكنوا من أخذ القرار المناسب في الوقت المناسب.

6. 1 العرض المرئي للبيانات

تتعدد طرق عرض البيانات عموما, والبيانات الناتجة عن عملية تحليل البيانات خصوصا. فمنها العرض النصي المدعم بالنتائج الرقمية ومنها جداول البيانات. ولكن أيا كانت الطريقة فمن المهم دعمها بعرض مرئي للبيانات المستخلصة. فكما يقال صورة خير من ألف كلمة. لذلك تضمنت التقارير غالبا أنواع مختلفة من الرسوم البيانية, ولعل من أكثرها شيوعا :

- الرسم البياني الدائري (Pie chart)

يستعمل أساسا لبيان النسب المئوية التي تمثلها مختلف قطع المخطط الدائري بصفة مختصرة تسمح بأخذ فكرة



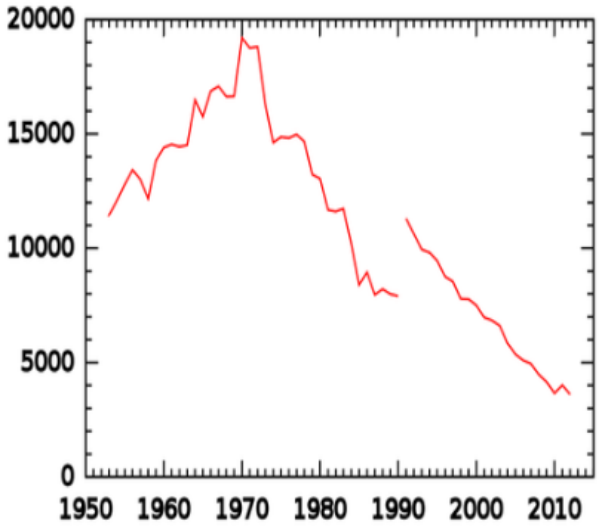
واضحة وشاملة عن البيانات.

رسم بياني دائري[17]

- الرسم البياني الخطي (Line graph)

هذا النوع من الرسوم يسمح أساسا بمتابعة تغير خاصية ما بتغير أخرى. تسمح مثلا بمتابعة سعر منتج ما على امتداد فترة زمنية. الخاصية هنا موضوع المتابعة هي السعر والمتغير هو الزمن.

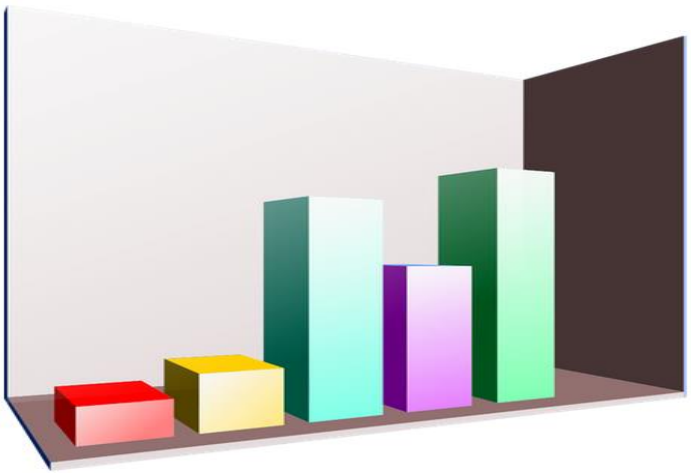
17 صورة مأخوذة من موقع https://publicdomainvectors.org



رسم بياني خطي[18]

- الرسم البياني الشريطي (Bar graph)

عندما يتعلق الأمر بمقارنة قيمتين متجاورتين حسب عدد من الخصائص موضوع التحليل يتم استعمال هذا النوع من الرسوم. ويتم وضع الأشرطة عموديًا أو أفقيًا جنبا إلى جنب.



رسم بياني شريطي[19]

- رسم بياني للنقاط المبعثرة (Scatter chart)

18 صورة مأخوذة من موقع https://publicdomainvectors.org

19 صورة مأخوذة من موقع https://publicdomainvectors.org

هذا النوع من الرسوم البيانية مفيد في عملية تحليل البيانات فهو يسمح بعرض الاحداثيات الرقمية على امتداد المحاور الأفقية والرأسية للبيانات. هذا من شأنه أن يتيح البحث عن المؤشرات والأنماط بين متغيرين.

هذه بعض الأمثلة عن الرسوم البيانية التي يتم اعتمادها في تحليل البيانات. كما أشرنا سابقا, الهدف منها إضافة لجعل البيانات متاحة وسهلة القراءة, تيسير عملية استثمارها من أجل استخراج العلاقات الكامنة بينها .

6. 2 الأساليب المرئية الحديثة لعرض البيانات

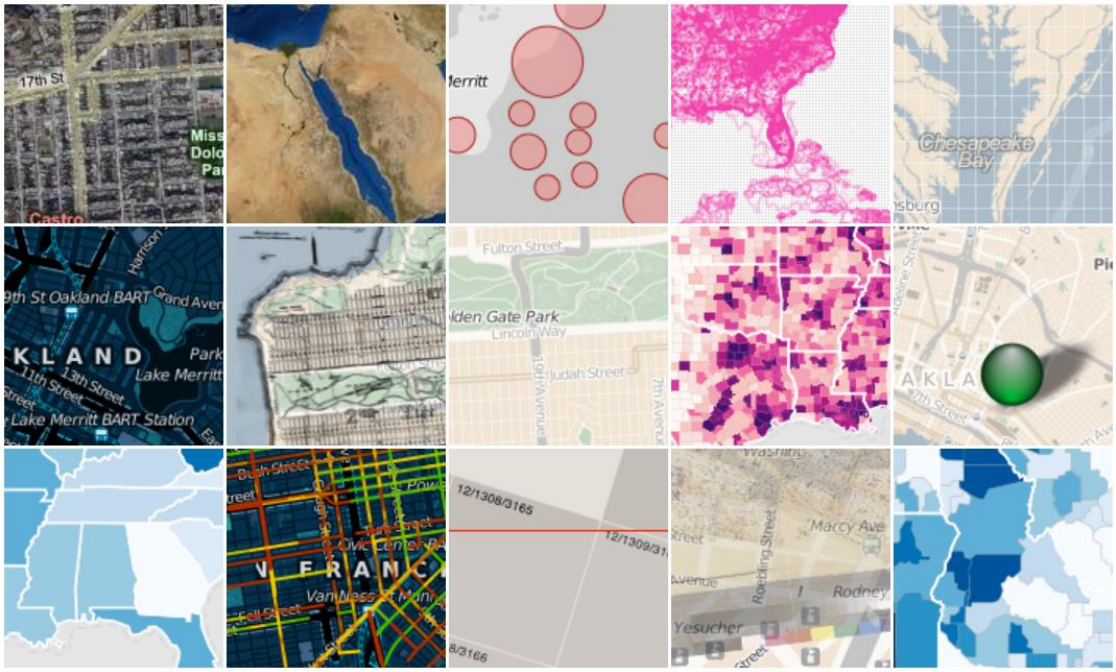
مع ظهور البيانات الضخمة ظهرت موجة جديدة من الأساليب المرئية الحديثة التي تسمح بمراجعة أسهل وأسرع لمخرجات عملية تحليل البيانات. بعض هذه الوسائل موجهة لعرض الخرائط, وأخرى موجهة لعرض النصوص, وصنف آخر موجه لعرض البيانات الرقمية, كما نجد تلك التي تسمح بعرض التصرفات والمشاعر, وأيضا تلك التي تعرض العلاقات والروابط.

- عرض الخرائط

الفكرة الأساسية من وراء استعمال عرض الخرائط تتمثل في ربط المعطيات بالمجال الجغرافي. لذلك تم توفير جملة من الأدوات التي تسمح بهذا الربط وبعرض الخرائط مع البيانات الخاصة بكل منطقة جغرافية.
للتمتع بهذه الميزة يجب استخدام أدوات مناسبة لهذه النوعية من العرض. فعلى سبيل المثال نجد PolyMaps وهي عبارة عن مكتبة برمجية معدة بلغة الجافا سكريبت (JavaScript) تسمح بربط البيانات بالخرائط مع توفير

جملة من الوظائف التي تسهل عملية سبر البيانات وقراءتها. وهي متوفرة عبر الرابط التالي:

http://polymaps.org



بعض الخرائط المنجزة ب PolyMaps[20]

الموقع الرسمي يعرض جملة من نماذج الاستخدام, ومنها عرض مدى انتشار استخدام الإنترنت على مستوى العالم, ومنها نسبة البطالة أو الكثافة السكانية حسب التوزيع الجغرافي, وغيرها من النماذج.

- عرض النصوص

إحدى وسائل عرض النصوص الحديثة والمبتكرة هي غيوم الكلمات (Word clouds). هذا الأسلوب يستعمل أساسا لتبيان الكلمات الأكثر تأثيرا والتي قد تعبر عن آراء أو عن مشاعر أو عن كلمات مفاتيح.

20 صورة مأخوذة من موقع http://polymaps..org

نموذج لغيوم الكلمات وكيفية إبرازها للكلمات الأكثر تأثيرا[21]

- عرض المعطيات

إحدى الطرق العصرية لعرض المعطيات والبيانات التحليلية هي استعمال ما يسمى بلوحة القيادة (Dashboard). تسمح لوحة القيادة بتجميع العديد من طرق العرض المرئي للبيانات بأنواعها مما يسمح للمستخدم بإلقاء نظرة سريعة على مختلف المؤشرات ومن ثمة يمكنه الاهتمام بمؤشر بعينه إذا شاء.

بناء لوحات القيادة يعتمد اليوم على برمجيات جاهزة متوفرة ومنها ما هو مجاني بل ومفتوح المصدر. كنموذج يمكن ذكر برنامج Mozaïk والمتوفر على الرابط التالي: http://mozaik.rocks

21 صورة مأخوذة من موقع https://publicdomainvectors.org

وهو برنامج بني اعتمادا على جملة من التكنولوجيات الحديثة والمجانية. ويسمح ببناء لوحة تحكم عصرية يمكن أن تحتوي على أنماط مختلفة من المؤشرات المرئية. هذه المؤشرات هي عبارة عن ملحقات (Extensions) متوفرة حسب الحاجة. كما يمكن برمجة وإضافة ملحقات جديدة حسب الحاجة.

مثال عن لوحة تحكم Mozaïk معروضة على الموقع الرسمي[22]

7. تعلم الآلة

يعتبر تعلم الآلة (Machine learning) أحد أهم وأكثر التكنولوجيات استخداما في مجال التحليل الآلي للبيانات وهو يندرج ضمن ما يعرف بالذكاء الاصطناعي (Artificial Intelligence). الهدف الأول وراء استعمال هذه التكنولوجيا هو تحويل المعلومات (Information) إلى معارف (knowledge).

22 صورة مأخوذة من موقع http://mozaik.rocks

عمليا يتعلق الأمر باستكشاف الأنماط (Patterns) المتخفية في البيانات الضخمة. إذ أن مع ازدهار هذه الأخيرة وإتاحة تخزينها ومعالجتها بأقل كلفة, أصبحت الحاجة إلى أتمتة تحليل البيانات ملحة أكثر من أي وقت مضى نظرا إلى صعوبة, بل استحالة التحليل الغير آلي لهذا الكم من البيانات.

إذا الفكرة الأساسية لتعلم الآلة تتمثل في اكتشاف الأنماط المتخفية في البيانات والتي تسمح بالحصول على المعارف اللازمة للإجابة على الأسئلة التي نود الإجابة عليها. هذه الأخيرة تتمحور حول إشكاليتين أساسيتين: توقع النتائج وأخذ القرارات. بالتالي مجال الاستخدام واسع ومتنوع بتنوع مجالات تحليل البيانات. ومن المعلوم أننا اليوم نستخدم الكثير من التطبيقات التي تعتمد على تعلم الآلة دون أن ندرك ذلك بالضرورة.

7. 1 نبذة حول المبادئ العامة لتعلم الآلة

يعتمد تعلم الآلة على نوع من الخوارزميات مختلف عن النوع السائد منها. فالخوارزميات السائدة تتمثل أساسا في جملة من الأوامر التي ترشد الحاسوب حول كيفية تحويل المعطيات إلى نتائج وذلك باتباع تلك الأوامر كل مرة بنفس الطريقة. أما خوارزميات تعلم الآلة فلا تقدم التمشي المفصل لتحويل البيانات إلى نتائج. إنما تقدم البيانات والنتائج المنتظرة وعلى الآلة أن تتعلم كيف تقوم بعملية التحويل وفق المنهج العلمي الذي تعتمده تلك الخوارزميات. لذلك تعمد الآلة إلى التعلم من خلال محاولات عدة لإيجاد الدالة الرياضية الأكثر ملاءمة قصد استنتاج النتيجة بناءا على المعطيات المقدمة مع اقل هامش خطأ ممكن وفي وقت معقول.

بالتالي تصبح الآلة قادرة على استنتاج الإجابة اعتمادا على البيانات دون الحاجة إلى تعليمات مسبقة. ذلك كما أشرنا أعلاه يتم من خلال تعلم الآلة من خلال تجربة القواعد (rules) التي تسمح باستنتاج النتيجة وقياس دقة الإجابات وهكذا حتى التوصل إلى مجموعة القواعد الأمثل للإجابة على الإشكال المطروح.

تجدر الإشارة أيضا إلى وجود عدة فئات من خوارزميات تعلم الآلة وهي:

- التعلم بالإشراف (Supervised learning)
- التعلم دون إشراف (Unsupervised learning)
- التعلم بشبه الإشراف (Semi-supervised learning)
- التعلم المعزز (Reinforcement learning)

كل هذه الأنواع تخضع لنفس مبدأ التعلم الآلي المذكورة أعلاه ولكنها تعتمد على مقاربات مختلفة لإيجاد القواعد التي تسمح بالإجابة على الإشكال المطروح.

2 .7 مراحل بناء نظم تعلم الآلة

عندما يتعلق الأمر بتحليل البيانات قصد الإجابة على إشكالية ما باستخدام تعلم الآلة, يجب بناء نظام معلوماتي. لبناء هذا الأخير يجب المرور بعدة مراحل. المرحلة الأولى تتعلق بجمع البيانات بما أن الآلة تتعلم من البيانات المتوفرة. وقد سبق وأن تناولنا مسألة جمع البيانات مع ذكر مصادرها وأنواعها وطرق تخزينها ومعالجتها.

ما يمكن إضافته هنا, هو أن مستخدم البيانات هو الآلة. وبما أن الآلة غير قادرة على التمييز, فإن مهندسي البيانات يعمدون إلى إعداد البيانات قبل تقديمها إلى الآلة. عملية الإعداد تتضمن تحويل البيانات الخام إلى الصيغة الأنسب للتحليل. من ضمن هذه الإعدادات تجميع البيانات المختلفة ودمجها لبناء مجموعة بيانات (Dataset) تأخذ شكل مصفوفة (Matrix). من ضمن الإعدادات أيضا استبدال البيانات الفئوية (Categorical) إلى بيانات رقمية (Numerical). هذا التمشي أساسي بما أن تعلم الآلة يعتمد على النمذجة الرياضية الرقمية. فعلى سبيل المثال لو كانت البيانات حول الطلبة, واحتوت عمودا خاصا باختصاص الطالب, فهذا الاختصاص هو من ضمن البيانات الفئوية. في هذه الحالة, أحد الحلول الممكنة هو إسناد رقم لكل اختصاص لتصبح البيانات رقمية خالصة.

كما تتضمن هذه المرحلة من الإعداد عملية تنظيف البيانات. هنا يتعلق الأمر أساسا بمعالجة البيانات الناقصة (missing data) والبيانات ذات القيم المتطرفة (outliers). هذه الأخيرة هي إما بيانات مغلوطة ناتجة عن خطأ في الرقن مثلا, أو استثناءات لا تمثل القاعدة. وتنظيف هذه البيانات يتم إما بالحذف أو باستبدالها بقيم لا تؤثر على النتيجة النهائية المطلوبة.

المرحلة الموالية تتعلق باختيار فئة تعلم الآلة. يبنى هذا الإختيار على عنصرين أساسيين هما:

- طبيعة البيانات المتوفرة (مثلا وجود القرائن أو عدم وجودها)
- الهدف المرجو من عملية تحليل البيانات (مثلا طبيعة القرينة التي نبحث عنها هل هي قيمة منفصلة أو قيمة متواصلة)

أمام توفر العديد من خوارزميات تعلم الآلة بالنسبة لكل فئة, فإن المرحلة التالية تتعلق باختيار الخوارزمية الأكثر دقة بالنسبة للبيانات التي تغذيها والقرائن التي يتم الحصول عليها. هذه المرحلة إذا هي عبارة عن عملية تقييم. وككل عملية تقييم فهي تعتمد على مؤشرات تسمح بقياس مدى دقة كل خوارزمية.

المرحلة الأخيرة, بعد اختيار خوارزمية بعينها, تتمثل عادة في ضبط متغيرات هذه الخوارزمية من أجل الحصول على دقة أكبر في النتائج (القرائن). ثم تتم عملية نشر التطبيقة ووضعها على ذمة المستخدمين. إذ يتم عادة الاعتماد على الحوسبة السحابية كبيئة نهائية لنشر التطبيقة. حيث توفر الكثير من المرونة والقدرة على التأقلم مع حجم البيانات وسرعة المعالجة المطلوبة.

7. 3 فئات خوارزميات تعلم الآلة

لقد سبق أن ذكرنا مختلف فئات خوارزميات تعلم الآلة وذكرنا اختصارا عوامل اختيار فئة بعينها عند بناء مشروع تحليل آلي للبيانات. كما ذكرنا أيضا أن في كل فئة توجد مجموعة من الخوارزميات المتعددة يتم الاختيار بينها اعتمادا على جودة المخرجات. لتبرير هذا التعدد والتنوع في الخوارزميات يتم الإشارة غالبا إلى نظرية تسمى "لا غذاء مجاني" (No free lunch theory) والتي تعني أنه لا توجد خوارزمية تعمل بكفاءة في كل الحالات والوضعيات الممكنة. في ما يلي سنتناول كل فئة لنبين أهم مميزاتها.

7. 3. 1 التعلم بالإشراف

يتم اختيار هذه الفئة من خوارزميات تعلم الآلة (Supervised learning) عندما يكون المنطلق هو مجموعة بيانات مصنفة (Labelled data). المقصود بالبيانات المصنفة تلك البيانات التي تحتوي في الآن نفسه (المدخلات) وأيضا نتائج التحليل المقترنة بها وهي القرائن (المخرجات).

على سبيل المثال لو فرضنا أن البيانات هي بيانات تعلم خاصة بالطلبة, وأن الهدف هو تصنيف الطلبة إلى فئتين: طلبة يعانون من صعوبات تعلم وطلبة لا يعانون من صعوبات تعلم . في هذه الحالة البيانات المصنفة هي تلك التي تحتوي تصنيف كل طالب, مما يسمح للآلة أن تتعلم كيف تقوم بالتصنيف من خلال الأمثلة التي تتمثل في المدخلات المقدمة والمخرجات الصحيحة المنتظرة. بعد مرحلة تعلم الآلة وتكوين نموذج يربط بين المدخلات والمخرجات, يصبح بالإمكان استعمال النموذج على بيانات جديدة (المدخلات) للتنبؤ بالقرائن الخاصة بها (المخرجات).

هذا في خصوص شرط إختيار فئة التعلم بالإشراف. لكن ضمن هذه الفئة نجد فئتين فرعيتين حسب نوعية القرينة التي يراد التنبؤ بها. حيث أن هنالك خوارزميات خاصة بالقرائن ذات القيمة المنفصلة (discrete) وتلك الخاصة بالقرائن ذات القيمة المتواصلة (continuous). فعلى سبيل المثال إذا كانت المعطيات تهم الطلبة وكان الهدف هو توقع انتماء الطالب لمجموعة الطلبة الذين يعانون من صعوبات التعلم أم لا, فإن هذه القرينة هي ذات قيمة منفصلة لأن هناك قيمتين فقط يمكن أن تأخذهما هذه القرينة. أما إذا كان الهدف هو توقع الدرجة النهائية للطالب, فهنا القرينة هي ذات قيمة متواصلة: فهي أي عدد عشري محصور بين القيمتين القصوى والدنيا.

الفئتين الفرعيتين لتعلم الآلة تحت الإشراف, هما خوارزميات التصنيف (Classification) وخوارزميات الانحدار (Regression). عندما يتعلق الأمر بتوقع قرينة ذات قيمة منفصلة يتم استخدام خوارزميات التصنيف. في حين عندما يتعلق الأمر يتوقع قرائن ذات قيمة متواصلة فيتم استخدام خوارزميات الانحدار.

2 .3 .7التعلم دون إشراف

هذا الصنف من خوارزميات تعلم الآلة (Unsupervised learning) يُستخدم أساسا عندما نفتقد البيانات المصنفة. هنا نتحدث عن بيانات غير مصنفة (Unlabelled data). الهدف من تحليل البيانات في هذه الحالة هو استكشاف الأنماط المتخفية والتي يصعب اكتشافها بطريقة أخرى داخل كمية ضخمة من البيانات الغير مصنفة مسبقا. لهذا نتحدث, عندما يتعلق الأمر بهذا الصنف من البيانات, عن اكتشاف المعارف (Knowledge discovery).

التعلم دون إشراف يساعد على القيام بما يسمى تحليل البيانات الاستكشافي (Exploratory data analysis) والذي يسمح بالقيام بعدة أصناف من المعالجات. فيما يلي أهمها:

- التجميع (Clustering)

وهي من أكثر الأصناف إستعمالا من ضمن أصناف خوارزميات تعلم الآلة دون إشراف. ويتعلق الأمر بتقسيم البيانات آليا إلى مجموعات اعتمادا على خصائص البيانات ومدى تقاربها.

فعلى سبيل المثال يمكن استعمال هذا الصنف عندما يتعلق الأمر ببيانات الطلبة. فتقوم الآلة بتقسيم الطلبة إلى مجموعات متجانسة اعتمادا على الخصائص التي تضمنتها البيانات المتوفرة. هذا يسمح باكتشاف العوامل المشتركة

لفئات من الطلبة مما يسمح من ناحية باكتشاف الخصائص التي تجمعهم وتسمح من ناحية أخرى بتوقع انتماء الطلبة الجدد لأحد هذه المجموعات. عادة المجموعات تحمل سيمياء واضحة, فقد تكون المجموعات تمثل على سبيل المثال الطلبة المتميزين والطلبة المتوسطين والطلبة ذوي مستوى دون المطلوب.

- الربط (Association)

هذا الصنف من خوارزميات التعلم دون إشراف يتم استخدامه عادة من أجل إستكشاف القواعد التي تصف البيانات من الناحية العلائقية. فلو عدنا لمثال بيانات الطلبة, فقد يسمح هذا الصنف من الخوارزميات بربط ندرة مشاركة الطالب في المنتديات الحوارية خلال التعلم عن بعد بالحصول على درجات ضعيفة. كما يمكن اكتشاف الرابط بين تميز الطالب في مادة ما بتميزه في مواد أخرى درسها مسبقا. وهذا يسمح بالقيام بتنبؤات خاصة بالطلبة الجدد اعتمادا على هذه الروابط المكتشفة.

- كشف عيوب في البيانات (Anomaly detection)

يقوم هذا الصنف من الخوارزميات بالتعرف إلى العناصر النادرة أو الغير معتادة, أي المختلفة عما هو سائد في البيانات موضوع المعالجة. أحد الأمثلة العملية والواسعة الإنتشار لاستعمال هذه الخوارزميات هو مجال السلامة, ومنها سلامة المعاملات البنكية.

فعلى سبيل المثال إذا كانت البيانات تمثل استعمال حريف لبطاقته البنكية, فإن اكتشاف استعمال غير معتاد ومختلف عما اعتاده الحريف قد تدل على عملية استعمال غير شرعي للبطاقة من قبل لص أو نتيجة لعملية قرصنة للبطاقة.

- اختزال الأبعاد (Dimensionality reduction)

البيانات هي مجموعة من القيم المسندة لكل خاصية (Feature) من الخصائص التي تصف موضوع التحليل. فلو أخذنا كمثال البيانات التي تصف استخدام الطلبة لمنصة تعليم عن بعد, فيمكن أن نجد من ضمن الخصائص عدد الساعات المقضاة في استعمال المنصة, عدد المشاركات في المنتديات, وغيرها من الخصائص. غير أن بعض هذه الخصائص قد تكون غير ذات جدوى لو أن الهدف هو التنبؤ مثلا بنجاح أو رسوب الطالبة. فعلى سبيل المثال قد تكون خصائص مثل الجنس أو العمر أو عنوان البريد الإلكتروني من الخصائص الغير مؤثرة في نتائج الطلبة.

إذا هذا هو الهدف من استخدام خوارزميات اختزال الأبعاد: الحفاظ فقط على خصائص البيانات (Data features) ذات الجدوى في تحليلها. من ثمة الاقتصار على هذه الخصائص سيسمح بتحسين دقة ونجاعة الخوارزميات المستخدمة لتحليل هذه البيانات.

7. 3. 3 شبه التعلم بإشراف

يتم استخدام هذا الصنف من تعلم الآلة (Semi-supervised learning) عندما يكون هنالك جزء يسير من البيانات المصنفة (Labelled) من ضمن عدد ضخم من البيانات الغير مصنفة (Unlabelled). الهدف هو عادة إما تقسيم البيانات (Clustering) أو اكتشاف الروابط (Association) أو غيرها من الاستعمالات التي سبق وأن تناولناها حين قدمنا خوارزميات التعلم دون إشراف. الإختلاف هنا هو أن وجود نزر يسير من البيانات المصنفة يسمح بتوجيه نتائج الخوارزميات من أجل تحسينها.

أحد الأمثلة التي يمكن أن نذكرها هنا كنموذج لاستخدام هذا الصنف من الخوارزميات, هو تحليل صور الأشعة الطبية من أجل الكشف عن الأورام الخبيثة. حيث أن تصنيف الصور ممكن نظريا من قبل الخبراء. لكن عمليا تصنيف عدد ضخم من الصور يكون مكلفا نظرا للحاجة إلى هؤلاء الخبراء وإلى وقت طويل لتحليل وتصنيف كل صورة. في هذه الحالة يمكن الاكتفاء بتصنيف عدد محدود منها ثم استعمال خوارزميات شبه التعلم بإشراف لتصنيف بقية الصور.

7. 3. 4 التعليم المعزز

خوارزميات التعلم المعزز(Reinforcement learning) تعمل بناءا على مبدأ المكافآت (Rewards). الفكرة تتمثل في في استعمال ردود إيجابية أو سلبية من أجل تحسين جودة آداء الخوارزمية. هذه الفكرة هي نفسها التي نستعملها نحن البشر في التعليم والتعلم في كثير من الأحيان, طبقا للنظرية السلوكية (Behaviorism).

هذه الخوارزميات شديدة التعقيد رغم بساطة المبدأ العام. توجد بعض الأمثلة الناجحة لتطبيقها, منها استخدامها من أجل التصرف في نظم التبريد لمراكز البيانات (Data centers). إذ تستهلك نظم التبريد الكثير من الطاقة الكهربائية المكلفة. هنا يتم استعمال خوارزميات التعلم المعزز من أجل الاقتصاد في الطاقة. حيث تتعلم الآلة من خلال التعزيز الإيجابي الذي يتم حين ينخفض استهلاك الطاقة مع المحافظة على درجة حرارة ضمن النطاق المسموح. في حين يتم التعزيز السلبي عندما يرتفع مستوى استهلاك الطاقة أو عندما تصبح درجة الحرارة غير مقبولة.

كما أن هنالك تجارب أخرى أثبتت فعالية هذه الخوارزميات منها قدرة الآلة على استنتاج قواعد واستراتيجيات اللعب دون الحصول على أي معطيات مسبقة ! حيث تم استعمال Google Deep Mind والتعلم المعزز لدراسة قدرة الآلة على تعلم لعب ألعاب الأتاري (Atari games) فقط من خلال مراقبة الشاشة وقراءة مجموع النقاط المتحصل عليها. وكانت النتائج مبهرة.

7. 4 تقييم خوارزميات تعلم الآلة

كنا قدمنا الأصناف الأربعة لخوارزميات تعلم الآلة وبينا متى تستخدم كل منها لتحليل البيانات مع تقديم أمثلة ملموسة. ولكن إذا اخترنا إحدى هذه الأصناف فالسؤال الموالي هو أي خوارزمية هي الأنسب للإستخدام في هذا الصنف؟ فكما أشرنا سلفا فإن نظرية "لا غذاء مجاني" (No free lunch theorem) تعني انتفاء قدرة أي خوارزمية على العمل بكفاءة مع كل الحالات الممكنة.

فمثلا لو كانت البيانات التي بحوزتنا مصنفة, أي أن لدينا القرينة المناسبة لكل عنصر في البيانات, فهذا يعني أننا سنستخدم خوارزميات تعلم الآلة الخاضع للإشراف (Supervised learning). ولنفترض مثلا أن المسألة تتعلق بالتصنيف (Classification), هنا نجد العديد من الخوارزميات التي تمكن من تحليل هذه البيانات بغرض تصنيفها. من ضمن هذه الخوارزميات يمكن أن نذكر:

- Logistic Regression
- K-Nearest Neighbors
- Support Vector Machine
- Naive Bayes
- Decision Tree Classification
- Random Forest Classification
- Gradient Boosting Classification

التمشي المتعارف عليه يتمثل في تجريب عدد من الخوارزميات ثم المقارنة بينها. هذه المقارنة تستوجب وجود مقاييس ومن ضمنها مصفوفة الارتباك (Confusion matrix). في هذه المصفوفة الأسطر تمثل مختلف الأصناف الحقيقية للبيانات والأعمدة تمثل الأصناف التي تنبأت بها الخوارزمية.

فلنأخذ مثال تصنيف الطلبة لثلاثة أصناف : متميز(A), متوسط (B) ودون المطلوب (C). البيانات في هذه المثال تشمل شريحة من 1000 طالب. فلنقل أن 80% من البيانات تم استخدامها لتعلم الآلة من خلال

خوارزمية Logistic Regression و ال20% الباقية تم استخدامها لتقييم الخوارزمية. النتائج المتحصل عليها يمكن تقديمها من خلال مصفوفة الارتباك التالية:

	A	**B**	**C**
A	47	2	1
B	8	70	2
C	2	3	65

الآن كيف يمكن فهم وتحليل محتوى البيانات التي تقدمها المصفوفة ؟

كما أشرنا أعلاه فإن كل سطر يمثل الصنف الحقيقي للبيانات. بالتالي فيمكن من خلال السطر الأول تعداد الطلبة المتميزين (صنف A) وهو 50 طالب (مجموع 47 مع 2 مع 1). أما الأعمدة فتشير إلى تنبؤات الآلة باستخدام خوارزمية Logistic Regression. حسب ما ورد في المصفوفة فإن من ضمن ال50 طالب من صنف A نجحت الآلة في التعرف إلى الصنف الصحيح في 47 حالة. وصنفت 2 من الصنف A باعتبارهم من الصنف B و 1 من الصنف A باعتباره من الصنف C. بنفس الطريقة يمكن أن نفهم ما ورد في السطرين التاليين.

عند تجربة خوارزمية أخرى من المهم استخدام نفس البيانات وتدريب الآلة بنفس الطريقة ثم تجربتها وأخيرا إصدار مصفوفة الإرتباك. فمثلا لنعتبر أن هذه المصفوفة هي نتيجة استخدام خوارزمية Random Forest Classification:

	A	**B**	**C**
A	49	1	0
B	4	75	1
C	3	7	60

هنا دائما السطر الأول يمثل الطلبة ال50 من الصنف A. الخوارزمية تعرفت بنجاح ل49 منهم كما يبينه العمود الأول. بينما فشلت في التعرف لحالة واحدة, حيث صنفتها من ضمن الصنف B.

قد يبدو من هذه المصفوفة أن خوارزمية التصنيف الثانية أفضل من الأولى. لكن وجب التنويه إلى أن قراءة الأرقام واستخلاص الأحكام يجب أن يتم بناءه على الهدف الأسمى من وراء تحليل بيانات الطلبة. فإذا كان الهدف الأسمى والأهم هو التعرف على الطلبة من الصنف C الذين يحتاجون إلى الدعم والإحاطة, فإن الخوارزمية الأولى تصبح الأفضل بما أنها الأدق في التعرف إلى هذا الصنف !

إضافة إلى مصفوفة الإرتباك, والتي تبين مدى ارتباك الآلة في تحديد الأصناف في مثالنا هذا, فهناك مقاييس أخرى يمكن اعتمادها ومنها:

- الصحة (Accuracy)
- الدقة (Precision)
- الحساسية (Sensitivity)
- النوعية (Specificity)
- مقياس F-1

هذا في خصوص مقاييس التقييم الخاصة بخوارزميات التصنيف (Classification). أما في خصوص خوارزميات الإنحدار (Regression) والخاصة بالتنبؤ بالقرائن ذات القيمة المتصلة (Continuous) فهنالك مقاييس أخرى. فعلى سبيل المثال إذا كانت البيانات تخص الطلبة, إذا القيمة المتصلة التي يمكن أن نسعى للتنبؤ بها يمكن أن تكون المجموع النهائي الذي يمكن أن يكون عدد عشري بين الصفر والمائة.

في حالة خوارزميات الإنحدار مقاييس التقييم المعتمدة تهدف إلى قياس مدى التباعد بين القيم الصحيحة والقيم التي تنبأت بها الخوارزمية. هنا ثلاث مقاييس عادة يتم الإعتماد عليها:

Mean Square Error (MSE)

Root Mean Square Error (RMSE)

Mean Absolute Error (MAE)

هذا لا ينفي وجود لمقاييس أخرى أقل إستخداما, منها على سبيل الذكر لا الحصر:

Max Error

Mean Absolute Error

Median Absolute Error

R^2 Regression Score Function

D^2 Regression Score Function

كل هذه المقاييس تمثل هامش الخطأ بالنسبة للخوارزمية, الهدف منها إذا هو مقارنة آداء الخوارزميات لاختيار أفضلها وأنسبها حسب الهدف من التنبؤ وحسب نوعية البيانات.

8. الخاتمة

تناولنا على امتداد هذا الباب مسألة تحليل البيانات. الهدف هو فهم أهمية هذا الصنف من معالجة البيانات ولفت النظر لما يوفره من إمكانيات هائلة من أجل تحويل البيانات إلى معرفة. هذه المعرفة تسمح بأخذ القرارات الصحيحة والسليمة والاستباقية.

علم تحليل البيانات يمثل أداة قوية للغاية بأيدي الفاعلين وأصحاب القرار في أي مجال من المجالات. بالتالي من المؤسف أن يتم تجاهلها عمدا أو سهوا.

تحليلات التعلم

1. مقدمة

مع ظهور تقنيات المعلومات والاتصالات الناشئة (البيانات الضخمة ، وذكاء الأعمال ، والذكاء الاصطناعي ، وتحليلات البيانات ، وبيئات التعلم الذكية ، وما إلى ذلك) ، استغلت العديد من المؤسسات التعليمية هذه التقنيات وعملت على تطوير أنظمة التعلم الخاصة بها وفقًا لذلك. في الواقع ، تقدم هذه المؤسسات مستوى آخر من الراحة لمتعلميها من خلال إزالة العقبات التي قد تعيقهم أثناء تعلمهم: لا مزيد من مشكلة الوقت، لا مزيد من مشكلة التوفر ،لا مزيد من مشكلة النقل، لا مزيد من المشاكل اللغوية، لا مزيد من المشكلة الثقافية، لا مزيد من المشاكل المالية، إلخ.

مع هذا الاتجاه ، تغيرت أدوات وأساليب التدريس وتطورت حيث أصبح المتعلم مركز الاهتمام. لاكتساب القدرات المعلنة في الدورة التدريبية، تقدم بيئات التعلم لكل متعلم واجهة مخصصة يمكنه من خلالها متابعة الدورات وفق وتيرته الخاصة وباستخدام الموارد المختلفة المتاحة له والنماذج التي يرغب فيها: ملفات (PDF, Word, PowerPoint) كبسولات رقمية (Moocs أو MicroMoocs) ، فيديوهات، ألعاب تعليمية ، عروض توضيحية ، محاكاة ، إلخ.

هذا الانتشار لأدوات التعلم سمح بحل العديد من المشكلات التنظيمية والتقنية. كما ادى أيضا إلى توليد كتلة ضخمة من البيانات لكل طالب. مع هذه الكتلة الكبيرة ، لم تتمكن معظم المؤسسات من الاستفادة منها الاستفادة القصوى للحصول على نتائج بناءة من حيث تحسين أداء المتعلمين أو جودة التدريس المقدم.

في كل هذا ، لا يزال السؤال الأساسي هو: كيف يمكن الاستفادة بشكل فعال من هذه الكتلة الكبيرة من البيانات لتحسين أداء المتعلم ونوعية التدريس؟

من المهم امتلاك المهارات اللازمة لتجميع وربط هذه البيانات للتمكُّن من تزويد أصحاب المصلحة بالوسائل والأدوات التي تسمح بتشخيص مسار كل طالب بشكل صحيح وفردي مع خيارات للتنبؤات التي يمكن أن تخمن مسبقًا, على سبيل المثال, ما إذا كان المتعلم سينجح أم سيرسب. مع تزاوج الذكاء الاصطناعي وتحليل البيانات ، وظهور التقنيات ذات الصلة ، جاء مفهوم حديث (في عام 2016) لتقديم الإجابات اللازمة, وهو تحليلات التعلم (.Learning Analytics)

في هذه الباب سنقدم هذا المفهوم, إذ أن الفهم هو المدخل الأمثل و الأنجع من أجل تبنيه لاحقا .يتكون هذا الباب من أربعة أجزاء: الأول سيقدم المفاهيم الأساسية والمصطلحات المرتبطة بتحليلات التعلم الاختصاصات المرتبطة بهذا المفهوم. أما الجزء الثاني فسوف يتطرق إلى مفهوم جمع البيانات و الجانب الأخلاقي وقضية خصوصية البيانات في تحليلات التعلم. بالنسبة للجزء الثالث ، فقد خصصناه لتقديم المستويات الأربعة لتحليلات التعلم (التحليلات التشخيصية ، التحليلات الوصفية ، التحليلات التنبؤية و الوصفة). سيخصص الجزء الأخير للتحليل التنبؤي الذي يعتمد أساسًا على الرسوم البيانية المختلفة.

.2المفاهيم الأساسية لتحليلات التعلم

2. 1 الدوافع

كما سبق وأشرنا في الباب الفارط (تحليل البيانات) أنه في ظل المنافسة الكبيرة تحاول الكثير من المؤسسات استخراج القيم أو الأنماط من البيانات ويتم استخدام تقنية الحوسبة السحابية (Cloud Computing)وعدد من الخوادم (Servers) لاستخراج المعارف.

في ناتفليكس (Netflix)، يتم استخدام مواقع بث الفيديو(video streaming websites)، والبيانات الخاصة بسلوك المشاهدة لدى المستخدمين، والأفلام التي يشاهدونها، وتقييماتهم (ما يعجبهم، وما لم يعجبهم) للتوصية لمشاهدة الفيديو التالي. في مواقع التجارة الإلكترونية مثل أمازون (Amazon) وبناءً على بيانات عمليات الشراء، يمكن للنظام أن يوصي بالمنتجات التي يمكن شراؤها. في المجال الطبي، يمكننا استخدام البيانات الضخمة للحمض النووي لعلاج العديد من الأمراض.

أما في مجال التعليم ومع استخدام الأدوات والتقنيات الرقمية, مثل التطبيقات التعليمية أو الموارد عبر الإنترنت أو أنظمة إدارة المكتبات أو نظم إدارة التعلم مثل Moodle[23] أو Blackboard[24] أو بوابات التقييم عبر الإنترنت، فإن كمية كبيرة من البيانات يتم إنشاؤها خلال الأنشطة التعليمية للمستخدم . هذا هو الدافع الأساسي لمختلف الأبحاث والإنجازات المتعلقة بها في مجال تحليل بيانات التعلم.

---

23 https://moodle.org

24 http://www.blackboard.com

2. 2 التعريف بتحليلات التعلم

باستخدام الأدوات والمنصات الرقمية للتعلم يسهل اليوم الوصول إلى بيانات التعلم مثل بيانات الطالب والتي تتضمن عادة:

- الخلفية الأكاديمية
- الملف الشخصي (الجنس والعمر)
- الأداء في اختبارات الفصل الدراسي وامتحانات الفصل الدراسي
- المشاركات في الدورة (الحضور التفاعل مع مختلف الأدوات التعليمية(

هذه البيانات لا تقدم الكثير إذا لم تخضع للتحليل. هنا يأتي دور تحليلات التعلم والتي يمكن تعريفها بأنها مجموعة من التقنيات والممارسات التي تسمح بتحويل البيانات التعليمية إلى معارف ذات قيمة مضافة.

تعتمد تحليلات التعلم على جمع البيانات وتحليلها من أجل الإجابة على أسئلة تهم تحسين آداء المتعلم أو تحسين طريقة التدريس أو تحسين الدرس.

حسب موقع الويب الخاص بمؤتمر تحليلات التعلم والمعرفة الذي عقد سنة 2011 بكندا (2011'LAK)، أن تحليلات التعلم هي قياس وجمع وتحليل وإعداد التقارير عن البيانات حول المتعلمين وسياقهم ، لغرض فهم التعلم وبيئته وتحسينهما. بشكل أوضح، تقوم تحليلات التعلم بجمع البيانات ، وتحليلها وإبلاغها إلى أصحاب المصلحة حول المتعلم وسياق التقارير الخاصة بها ، من أجل تحسين تجربة وأداء المتعلمين أو توفير المحتوى المتكيف(Adaptive content) أو تحسين استراتيجية التدريس والبيئة التي يعمل فيها الطلاب.

هناك العديد من التعريفات تحليلات التعلم التي لا تزال قابلة للنقاش لأن هذا المجال يعتبر حديث العهد. لكن، عندما نتحدث عن تحليلات التعلم ، يجب علينا دائمًا الرجوع إلى مجتمع تحليلات التعلم والمعرفة (: SOLAR Society of Learning Analytics and Knowledge) الذي وفي فترة وجيزة استطاع أن يخلق ديناميكية بإرساء العديد من الانشطة العلمية مثل : انشاء مجتمع أبحاث تحليلات التعلم والذي يمكن زيارته على الرابط التالي: www.solaresearch.org. كما قام بإنشاء ثلاثة أحداث رئيسية مثل المؤتمر (Conference LAK) الذي يقام سنويا و مجلة (Journal LAK) وهي مجلة تنشر أبحاثا حول تحليلات التعلم وأيضًا المدرسة الصيفية لتحليلات التعلم (Summer School of Learning Analytics). ولهذا السبب ، إذا أردنا التعمق في تحليلات التعلم ، فيجب علينا مراجعة ما قام به هذا المجتمع العلمي بصفته عملا مرجعيا.

الغرض الأكثر انتشارا لاستعمال تحليلات التعلم هو فهم تجربة المتعلمين بغرض تحسينها. بمعنى آخر، الغرض هو جمع البيانات المتوفرة) بيانات تفاعل الطالب مع التطبيقات والإعدادات التعليمية المختلفة، (... ثم القيام بتحليل تلك البيانات للخروج باستنتاجات وتوصيات من شأنها تحسين فعالية التعلم وجودة التدريس. فيمكن بالتالي فهم سلوك الطلاب من خلال تطبيق أساليب تعلم الآلة أو الأساليب الإحصائية مثل التنقيب عن البيانات التعليمية.

2. 3 مستخدمو تحليلات التعلم

يتعدد مستخدمو تحليلات البيانات. حيث أن أصحاب المصلحة في استخدامها هم أساسا وليس حصرا:

- المدرسين (المعلمين والأساتذة)،
- المتعلمين (الطلاب والتلاميذ)،
- المؤسسة التعليمية أو المنصة الإلكترونية التعليمية
- المدير أو الرئيس أو المشرف

وتختلف استخدامات تحليلات التعلم حسب نوعية الطرف المستفيد. فإذا كان أصحاب المصلحة هم من صنف المدرسين، تمكنهم تحليلات التعلم من تقييم أداء المتعلمين في الوقت الفعلي (Real time). على سبيل المثال ، إذا كان لدى المعلم لوحة معلومات للطلاب الذين يقومون باختبار معين أو يعملون في بيئة تعليمية ما، فيمكنه الحصول على نظرة ثاقبة في الوقت الفعلي حول أداء المتعلمين خاصة أولئك الذين يعرفون صعوبات. بالتالي يعرف المعلم أي طالب تحديدا يحتاج إلى مساعدة. كما يمكن للمدرس أن يذهب أبعد من ذلك ويقوم بالخطوات اللازمة لمساعدتهم. في حالات أخرى قد يسمح ذلك أيضا باكتشاف أن الفصل بأكمله لديه صعوبات في خصوص موضوع معين. بالتالي قد يتعين عليه تدريس ذاك الموضوع بطريقة أخرى أفضل وأكثر نجاعة.

المتعلم أيضا يمكن أن يستفيد من تحليلات التعلم. إذ يمكن أن يتلقى الطلاب معلومات حول أدائهم مقارنة بأقرانهم في الفصول الدراسية مما يساعدهم على المزيد من التحفيز لبلوغ أهدافهم وتحقيقها. أولا قد تسمح نتائج التحليلات للمتعلمين بالتنافس للحصول على أفضل النتائج من خلال مقارنة نتائجهم ومدى تقدمهم في الدرس ومدى انجازهم للفروض مع إما بقية المتعلمين عموما, وإما مع العشرة الأوائل, وإما مع صنف محدد من المتعلمين. الفكرة هنا هي استخدام تحليلات التعلم كوسيلة تحفيز.

ثانيا يمكن استخدام تحليلات التعلم كوسيلة تطمين للمتعلمين وضمان استمرارهم في عملية التعلم. فعلى سبيل المثال ، إذا أجاب المتعلم على سؤال بشكل خاطئ, ثم تبين أن في الفصل قرابة 40 بالمائة من الطلاب مخطئون أيضا في الإجابة على نفس السؤال, سيشعر الطالب أنه ليس الوحيد الذي أعطى إجابة خاطئة. بالتالي يتم تحفيز المتعلم على الاستمرار. فإنه إذا اعتقد المتعلم أنه الوحيد الذي يجيب خطئا فقد يصاب بخيبة أمل يمكنها ان تؤدي الى عدم اهتمامه بالاستمرار، والنتيجة النهائية ستكون حتما غير مرضية.

ثالثا تتيح تحليلات التعلم للمتعلمين أيضًا معرفة مدى تقدمهم في دورة تدريبية معينة. إذا علم المتعلم أنه يتقدم بنسبة 40 في المائة من المحتوى في هذه الدورة فهذا قد يسمح له بتثمين الجهد الذي بذله وتحفيزه لبذل المزيد من الجهد لمواصلة هذه الدورة. كما تمكن مثل هذه البيانات المتعلمين من مقارنة أدائهم مع أقرانهم كمحفز إضافي.

أخيرًا ، تعتبر تحليلات التعلم ذات أهمية كبيرة لأي مؤسسة تعليمية ، والتي يجب أن تقاوم في مواجهة المنافسة الحادة مع المؤسسات الأخرى. نتيجة لذلك ، يمكن أن تحقق تحليلات التعلم قيمة مضافة على عدة مستويات: إدارة الموارد البشرية ، وإدارة الموارد التعليمية ، وتقليل المخاطر ، وتحسين جودة التدريس ، والتخطيط الاستراتيجي. سيسمح هذا الخلق للقيمة للمؤسسة المعنية بوضع نفسها بقوة في السوق وأن يكون لها حوكمة رشيدة على نظام التعليم بأكمله. للقيام بذلك، يسعى القادة وصانعو السياسات جاهدين لفهم معنى وأهمية تحليلات التعلم وكيفية الاستفادة من تأثيرها لتحسين آداء الطلاب وجودة التعليم وبالتالي ترشيد الحوكمة. في هذا الصدد يواجه هؤلاء المسؤولين فرصًا وتحديات مختلفة أمام ظهور تحليلات التعلم. نظرًا لأنه مفهومها حديث وغير ناضج . فمن الطبيعي أن يكون لدى هؤلاء المسؤولين العديد من الأسئلة ليطرحوها .نذكر على سبيل المثال :

-كيف يمكن لأعضاء هيئة التدريس أن يصبحوا ممارسين بارزين في مجال تحليل التعلم ، وبناء خبراتهم ورفع مستوى ممارسة تحليل التعلم ، عبر الدورات والتخصصات ، بين أقرانهم وعبر مؤسستهم؟

-كيف يمكن لأعضاء هيئة التدريس والموظفين المؤسسين المشاركين في مبادرات نجاح الطلاب تضمين تحليلات التعلم لدعم التدخلات والإجراءات الديناميكية؟

-كيف يمكن لقادة المؤسسات وصانعي السياسات تطوير سياسات وممارسات داعمة وأدوات إدارة التعلم / الدورة التدريبية التي تطلق العنان للقوة التحويلية لتحليلات التعلم عبر مؤسساتهم وخارجها؟

-كيف يمكن لأصحاب العمل وصانعي السياسات توفير روابط فعالة للمعرفة المهنية والقوى العاملة التي من شأنها تعزيز وتوسيع نطاق نجاح الطلاب ليشمل العناصر الأكاديمية وتطوير المناهج الدراسية وقابلية التوظيف والوظيفة؟

ستكون الإجابات على هذه الأسئلة نتيجة لسلسلة من الإجراءات التي يتعين تنفيذها داخل المؤسسة التعليمية كجزء من مشروع تحليلات التعلم :

-ترسيخ تحليلات التعلم في النظم المؤسسية والممارسات واستراتيجيات نجاح الطلاب.

-تغيير أبعاد الثقافة التنظيمية حيث تشير بعض التجارب مع العديد من المؤسسات إلى أنه حتى عندما يتم إطلاق المبادرات بشكل مدروس ، فإن الثقافة التنظيمية لا تتغير بسرعة ، أو لا تتغير كلها مرة واحدة. علاوة على ذلك ، فإن فهم أبعاد التغيير المطلوب يمكن أن يتغير بمرور الوقت ؛ بالإضافة إلى ذلك ، يجب على اللجان المؤسسية ، من خلال التجربة والتفكير ، تطوير فهم أفضل لتدخلات نجاح الطلاب المدعومة بالتحليلات في الممارسة العملية.

-تغيير السياق التنظيمي والقدرة التنظيمية : لا يتعلق الأمر بالثقافة فقط ؛ يتعلق الأمر بجميع جوانب القدرة التنظيمية للتحليلات. يصور الجدول 26.2 الأبعاد الخمسة للقدرة التنظيمية ، التي تم تقييمها لمؤسسة عينة. وجدت الفرق المؤسسية التي تتولى تعزيز نجاح الطلاب أو تحسينه وجعله أولوية مؤسسية أنه من المفيد تقييم الوضع

الحالي لتنمية القدرات (يُطلق عليه أيضًا مؤشر الجاهزية / النضج) ، ثم تحديد الأهداف اللازمة لتعزيز نجاح الطلاب عبر التخطيط المعقول على فترة كافية.

يوضح الجدول أدناه التقييم الحالي، والنتيجة المستهدفة في خلال فترة خمس سنوات على سبيل المثال، و "الفجوة" بين الاثنين التي يجب إغلاقها بمرور الوقت (الأشرطة السوداء في عمود "النتيجة المستهدفة"). يحتاج قادة المؤسسات إلى التركيز على تحديد أهداف ممتدة لقدرة الطلاب على النجاح والتي من شأنها أن تضعهم في مكانة تمكنهم من تحقيق أهدافهم.

السياق المؤسسي / القدرة على التحليلات ونجاح الطلاب ، الحالي والمستهدف		
أبعاد السياق التنظيمي والقدرة التنظيمية	التقييم الحالي	النتيجة المستهدفة
القيادة		
تلتزم الإدارة العليا بتعزيز نجاح الطلاب وتعتبر تحليلات التعلم التنبؤية ضرورية لنجاح الطالب	4/5	5/5
التحليلات لديها مسؤول رفيع المستوى يمكنه إزالة الحواجز ودعم التمويل	3/5	5/5
تلتزم القيادة العليا بممارسة صنع القرار القائم على الأدلة وممارستها باستمرار	2/5	4/5
تم توفير التمويل والاستثمار المناسبين في التحليلات	3/5	5/5
الثقافة / السلوك		
تفضل ثقافة المؤسسة الأدلة القائمة على الأداء لاتخاذ القرار	2/5	3/5
تدرك ثقافة المؤسسة "ضرورة المعرفة" وممارسة "التحليلات العملية"	2/5	4/5
يقيم تعلم الطلاب وابتكارات النجاح وتأخذ الابتكارات الناجحة على نطاق واسع	2/5	5/5
التاكيد على التعاون في نجاح الطلاب ونجعل نجاحهم وظيفة الجميع	2/5	3/5
يحدد ويدمج نجاح الطالب ليشمل تطوير المناهج والمناهج الدراسية والعمل وتنمية المواهب	2/5	3/5
السياسات والعمليات والممارسات		
السياسات المؤسسية والإشراف على البيانات تفي بالقوانين الفيدرالية والولائية والمحلية	2/5	3/5
يتم توثيق سير العمل لعمليات نجاح الطلاب بشكل جيد	3/5	4/5
توجيه فرق التحالف ومتعددة التخصصات لنجاح الطلاب	2/5	4/5
التخطيط المتكامل وتوفير الموارد والتنفيذ والاتصال لنجاح الطلاب	3/5	4/5
تنمية المهارات والمواهب		
مهارات الابتكار / التعاون في نجاح الطالب	2/5	4/5
مهارات تحليلات التعلم المحددة: علم البيانات - تحليل البيانات وتفسيرها والتصور	2/5	4/5
يتم تطوير القدرة على إعادة اختراع عمليات دورة حياة الطالب (من البداية إلى النهاية) بشكل جيد	2/5	3/5
القيم الممكنة للتقييم والنتيجة: 1. استهلالي. 2. ناشئ ؛ 3. فعال . 4. فعال للغاية. 5. مثالي		
المصدر : Handbook of Learning Analytics", p.313, ed. SOLAR, 2017"		

2. 4 تحليلات التعلم و التنقيب عن البيانات التعليمية

هنا سنقدم مزيدًا من التوضيح حول مصطلحات تحليلات التعلم (LA : Learning Analytics) ، والتنقيب عن البيانات التعليمية (EDM :Educational Data Mining) والتحليلات الأكاديمية (AA :Academic Analytics) من أجل إزالة أي غموض أو خلط بينها.

في البداية سنهتم بالتنقيب عن البيانات.(Data Mining) هي عملية تحويل البيانات الخام إلى معلومات مفيدة : هناك الكثير من البيانات في وسائل التواصل الاجتماعي ، في التجارة الإلكترونية، في الصحة، في القطاع البنكي وفي التعليم. السؤال الذي يطرح هو كيف يمكن استغلال هذا الكم من البيانات على أحسن وجه؟ أي كيف يمكن استخدام هذا القدر من البيانات وتحويلها إلى معلومات مفيدة حتى تتمكن المؤسسة المعنية من تقديم توصيات أو تقديم إعلانات مفيدة أو اتخاذ قرارات فعالة. من بين التقنيات التي نستخدمها لتحليل البيانات تلك القائمة على مفهوم "التنقيب في البيانات".

يتم التنقيب عن البيانات، باستخدام تعدين النمط (Pattern Mining) أو تعدين العمليات (Process Mining) أو التنبؤ بشيء ما (Predicting) أوعن طريق القيام بالارتباطات والانحدارات (Correlations & Regressions) أو تطوير بعض خوارزميات التنبؤ. على سبيل المثال، يسمح لنا هذا النوع من الأدوات بالتنبؤ بالمنتج الذي سيشتريه الشخص، أو الفيلم الذي يرغب في مشاهدته. يستخدم التنقيب عن البيانات أيضًا في التسويق والتجارة الإلكترونية: فهو يدعم العثور على المنتج المراد شراؤه أو يوفر إعلانات بناءً على سلوك البحث الخاص بالمستخدم.

كذلك ، يمكن استخدام سلوك المستخدم في المتصفح أو محتوى البريد الإلكتروني لتقديم إعلانات مستهدفة أو إدارة مخاطر الائتمان في المجال البنكي . على سبيل المثال، بناءً على السلوك السابق للمستخدم، أو مقدار القرض الذي حصل عليه، أو ما إذا كان المستخدم قد سدد القرض في الوقت المحدد أم لا، يمكن استخدام كل هذه

البيانات لتقييم التصنيف الائتماني للمستخدم، ومن ثم ما إذا كان ينبغي تقديم قرض جديد أم لا. لذلك فان استخدام التنقيب عن البيانات يساعد البنوك على اتخاذ القرارات الصائبة في الوقت المناسب.

في مجال التعليم، يمكن أن يكون التنقيب عن البيانات مفيدًا في التحليل و وخاصة في كيفية تفاعل الطلاب في منتدى المناقشة: على سبيل المثال في حالة مساق هائل مفتوح على الإنترنتMOOC))، يمكن أن نجد آلاف المستخدمين المسجلين. البعض منهم سيتناقشون في المنتدى وسوف يولدون كمية كبيرة من البيانات التي إن تمت معالجتها يمكن من خلالها تقديم ملاحظات قيمة للمتعلمين. معالجة ذلك الكم من البيانات يمكن القيام به عن طريق التنقيب عن البيانات وذلك من خلال إيجاد الأنماط المفيدة في منتديات المناقشة. هنا لا بد من لفت النظر أنه عندما نتحدث عن كيفية استكشاف البيانات من MOOC أو مصادر التعلم الأخرى فقد انتقلنا هنا إلى مجال تطبيق خاص وهو التنقيب عن البيانات في التعلم (EDM – Educational Data Mining).

يشير هذا التنقيب إلى العملية المصممة (Designed process) لتحليل البيانات من بيئة المتعلم لفهم الطلاب بشكل أفضل في البيئة التي يدرسون فيها. ويركز EDM في الواقع على تطوير الأساليب والأدوات للبيانات من الإعدادات التعليمية (Educational settings) ، في حين أن تحليلات التعلم تحاول فهم سلوك الطالب في محاولة لنمذجة الطلاب من أجل تحسين أداء الطالب، وكذلك لتحسين استراتيجية التدريس.

بالنسبة للمؤسسات التعليمية، إذا تحدثنا عن EDM فهذا يعني استخدام بيانات حول استغلال الطلاب للموارد : كم عدد الطلاب الذين استخدموا الموارد مثل الكتب المتاحة أو نظم إدارة التعلم (LMS) مثل

Moodle[25]، هل قام الطالب بتحميل المهام في الوقت المحدد، أو المشاركة في الفصل. يمكن جمع كل هذه البيانات وتطبيق التنقيب عليها كما يمكن أيضا نمذجة سلوك تفاعل الطالب مع الموارد التعليمية. في هاته الحالة يمكن للمعهد اتخاذ قرار بشأن ما إذا كان ينبغي إستثمار المزيد على هذا المورد أم لا.

لذلك، فإن التنقيب عن البيانات في التعلم (EDM) يعني تطبيق التنقيب عن البيانات التعليمية والذي يمكن أن يكون مفيدًا في تطوير نموذج المتعلم، أو المساعدة في اتخاذ بعض القرارات على مستويات مختلفة. بينما لتحليل وتفسير هذا النوع من البيانات واستخراج المعارف المتخفية وراءها ، يجب استخدام مفهوم آخر وهو التحليلات الأكاديمية (AA :Academic Analytics).

حاليًا ، يُستخدم مصطلح ذكاء الأعمال(BI:Business Intelligence) لتقديم المؤشرات التي يمكن أن تساعد المديرين في اتخاذ قراراتهم على مستوى شركتهم ، بغض النظر عن المجال: الاقتصادي ، المالي ، الصحي ، الزراعي ، الصناعي ، إلخ. ولكن ، عندما تصبح المنظمة مؤسسة تعليمية ، يتم استبدال هذا المصطلح بالتحليلات الأكاديمية (AA :Academic Analytics).

يمكن للمسؤولين وصناع القرار استخدام هذه البيانات للتنبؤ بسلامة العمل بمؤسسة تعليمية ما .بل من الممكن الذهاب أبعد في التحليلات من أجل الإجابة على أسئلة من قبل: ما هي المؤسسة التعليمية التي لا تحقق أداءً جيدًا؟ ولماذا لا تعمل بشكل جيد؟ وما هي العوامل التي تؤثر على أدائها؟

يمكن النظر إلى هذا النوع من المعلومات على أنه تحليلات أكاديمية أو ذكاء أعمال. فعلى سبيل المثال، لنأخذ المعطيات التالية:

25 https://moodle.org

- حضور المعلمين،
- نسبة النجاح في دورة معينة،
- أداء مدرسة ما
- معدل تخرج الطلاب في جامعة معينة.

السؤال هو: أي منها ليس من ضمن نتائج التحليلات أكاديمية؟

الإجابة هي أن النسبة المئوية للنجاح في الدورة ما بعينها ليست من ضمن ما يتم استخراجه من خلال التحليلات الأكاديمية. في حين ترتبط جميع الخيارات الأخرى بالتحليلات الأكاديمية مثل حضور المعلمين في الجامعة بأكملها أو أداء مدرسة ما أو معدل تخرج الطلاب في الكلية. لكن النسبة المئوية للنجاح في دورة تدريبية مرتبطة بدورة معينة مرتبطة فقط بمسألة التعلم. لذا ، فهذه ليست تحليلات أكاديمية مقارنة بالخيارات الثلاثة الأخرى.

هذا بالنسبة التحليلات الأكاديمية (Academic Analytics), والتي تهدف كما أشرنا أعلاه إلى تحليل البيانات التي تهم مؤسسة تعليمية في مجملها. في حين أن تحليلات التعليم (Learning Analytics) تهتم بتحليل بيانات التعلم على مستوى أدق على غرار: القسم, الدرس, المتعلم, المعلم, المحتوى التعليمي.

هذا الجدول أسفله يختزل أهم الفوارق بين التحليلات الأكاديمية وتحليلات التعلم:

التحليلات الأكاديمية	تحليلات التعلم

تقديم الدعم لاتخاذ القرارات التشغيلية والمالية	تقديم الدعم لتحقيق هدف تعليمي محدد
التركيز على أعمال المؤسسة	التركيز على تعلم الطلاب
أصحاب المصلحة: المستويات الإدارية والتنفيذية	أصحاب المصلحة: المدربين والمتعلمين

جدول مقارنة التحليلات الأكاديمية وتحليلات التعلم

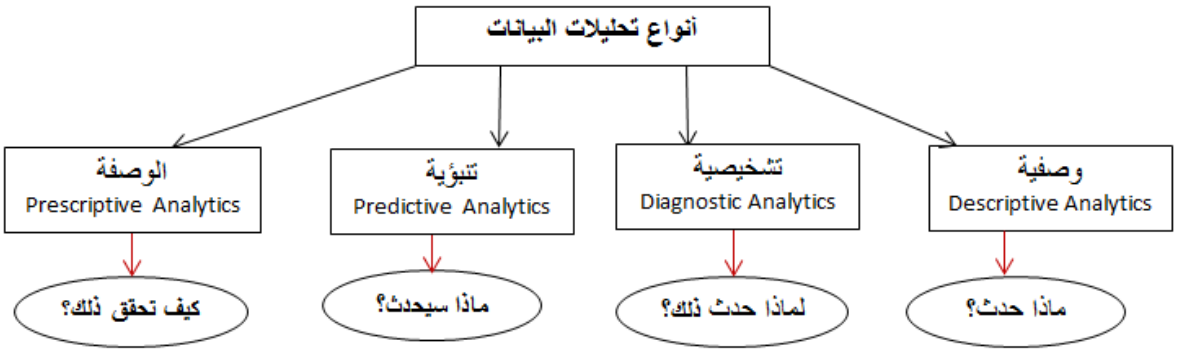
من هنا فصاعدا لن نتحدث عن التحليلات الأكاديمية لكننا سنتحدث عن تحليلات التعلم. سيتم شرح كيفية تطبيق تطوير النماذج للمتعلمين وأيضًا كيفية فهم أداء المتعلمين وتحسينه وكذلك مشاركة البيانات مع المعلمين.

2. 5 المستويات الأربعة لتحليلات التعلم

ذكرنا في الأقسام الأخيرة أنه يمكن تعريف تحليلات التعلم على أنها تطبيق لتحليل البيانات على نوع بعينه وهو بيانات التعلم. وتنقسم هذه التحليلات إلى أربعة أنواع : وصفية، تشخيصية، تنبؤية وتعليمية .

التحليل الوصفي يسمح بالإجابة على التساؤل التالي: ماذا حدث؟ لفهم ما حدث أثناء العملية التعليمية فإنه يتم إصدار وصف في شكل جداول أو مخططات أو كلمات مكتوبة أو بعض الرسوم البيانية وأي شيء من هذا القبيل.

أما التحليل التشخيصي للبيانات فيسمح بالإجابة على التساؤل التالي: لماذا حدث هذا؟ هنا أيضا يتم استخدام مختلف الوسائل البصرية والأشكال التوضيحية للإجابة عن هذا السؤال. فإذا فهمنا سبب حدوث ذلك، فالسؤال الموالي المنطقي هو: ماذا سيحدث بعد ذلك؟ هنا يأتي دور التحليلات التنبؤية. وإذا أصبح لدينا تصور واضح لما سيحدث بعد ذلك، فالسؤال الموالي الممكن هو: هل يمكننا تغيير نتيجة المتوقعة؟ هنا يأتي دور التحليلات التي تقدم الوصفة.



المستويات الأربعة لتحليلات التعلم

من خلال هذا التقديم قد نتبين أن هنالك تمشي منهجي في تطبيق مختلف مستويات التحليل. هذا التمشي ليس ضروريا, بل يمكن حسب الهدف المنشود الاقتصار على مستوى محدد. لكن يجب التفطن إلى أن المنهجية التي قدمناها والتمشي الذي يربط بين مختلف المستويات ينطبق على ممارسات عدة نراها في حياتنا اليومية.

فمثلا إذا ذهب مريض إلى الطبيب، فهذا الأخير سيسأله بداية: ماذا حدث لك؟هنا يقوم المريض بوصف أعراضه. من خلال فهم الوصف وبناءً عليه يقوم الطبيب بتشخيص الحالة فيسأل أسئلة أخرى على غرار: لماذا حدث ذلك؟ مثال : هل أكل المريض شيئًا خاصا الليلة الماضية؟ هل تناول الحليب مع القهوة؟ هل استهلك بعض الطعام المقلي؟ إذا، بناءً على هذه المعلومات ، يقارن الأعراض بمعلومات أخرى ربما تكون قد تكون متوفرة، ويحاول أن يأتي بنموذج للإجابة عن السؤال التالي: لماذا حدث لك ذلك؟ من هنا يمكن للطبيب التنبؤ بتطور حالة المريض في قادم الساعات والأيام المضاعفات الصحية التي يمكن أن يعاني منها إذا لم يتم علاجه. هذا هو التحليل التنبؤي. ثم بناءً على ذلك ، سيقدم الطبيب الوصفة الطبية التي من شأنها أن تسمح للمريض تلافي المضاعفات الصحية والتعافي بسرعة.

إذا بشكل عام عند زيارة الطبيب، نقوم أولاً بوصف الأعراض، ثم يقوم الطبيب بتشخيصها ، ثم يتنبأ هذا الأخير بما كان سيحدث. ثم يقدم العلاج الذي يعتقد أنه الأنسب. ما نود أن نشير إليه هو أنه, في بعض الحالات، قد يحدث خطأ في التنبؤ .لهذا قد يحتاج المريض الى اختبارات اخرى وتحاليل وصور أشعة لتحديد السبب الدقيق. هذا ما يحدث أيضا في تحليلات البيانات عموما وتحليلات التعلم خصوصا.

إذا المستوى الأول هو التحليلات الوصفية. هي تحليلات تصف أو تلخص البيانات الأولية وتجعلها قابلة للتفسير. يجب أن تمثل أهم المعطيات وتوفر الحقائق وملخص البيانات في حوامل سهلة القراءة والاستعمال. هذه التحليلات التي يجب أن تكون يسيرة الفهم من قبل مختلف أصحاب المصلحة. في هذه المرحلة قد تكون التحليلات الزمنية للبيانات مفيدة.

فلنأخذ مثال لطلبة الهندسة في اختصاص علوم الحاسب في إحدى مدارس المهندسين. يختار الطلبة, إضافة للمواد المشتركة, اختصاصا يسمح لهم بالحصول على مواد خاصة بالاختصاص. الاختصاصين هما هندسة الشبكات و هندسة البيانات. من ضمن المواد المشتركة, والتي تدرس لكلا الفريقين, مادة التجارة الإلكترونية. الأستاذ المشرف قرر القيام بتحليل وصفي للبيانات المتاحة حول هذه المادة للخمس السنوات الأخيرة ليرى إن كانت هنالك فوارق في الأعداد بين الطلبة حسب عدة متغيرات (عدد الغيابات, الجنس, ...). من ضمن التحليلات رسم بياني شريطي يبين معدل الأعداد المتحصل عليها من كل فريق للسنوات الخمس الأخيرة. هذا الرسم يبين أن معدل أعداد فريق هندسة البيانات كان دائما أقل وبصفة ملحوظة وثابتة عن معدل أعداد فريق هندسة الشبكات ! هنا نرى بوضوح أهمية التحليل الوصفي, الذي يسمح لاكتشاف مواطن الخلل الممكنة.

الأستاذ المشرف على مادة التجارة الالكترونية قرر أن يشخص الأسباب الكامنة وراء هذا الفرق الملحوظ في الأعداد. من الأسئلة التي خطرت على ذهنه وأراد أن يقوم بالتحاليل اللازمة لإثباتها أو نفيها هو مدى تأثير مواد الاختصاص إيجابا أو سلبا على أعداد مادة التجارة الإلكترونية.

هنا نمر إلا التحليل التشخيصي. فمن ضمن التحليلات التي قام بها المدرس هو دراسة مدى ارتباط (Correlation) الأعداد المسندة في مادة التجارة الإلكترونية مع أعداد كل مادة من مواد الاختصاص لكل فريق. استعمل المدرس مصفوفة الارتباط (Correlation Matrix) لدراسة الترابط الممكن بين الأعداد. هنا لاحظ المدرس أن أعداد مادة سلامة الشبكات, التي يدرسها فقط طلبة اختصاص هندسة الشبكات, لديها معدل ارتباط يصل إلى 80% بأعداد التجارة الإلكترونية. بالتالي فالتشخيص هو: يجب دراسة سلامة الشبكات من أجل فهم أفضل لمادة التجارة الالكترونية وبالتالي الحصول على أعداد أفضل.

هنا السؤال الموالي المطروح من قبل المدرس, هو ما هي النتائج المتوقعة للطلبة الجدد اختصاص هندسة البيانات مقارنة بنظرائهم اختصاص هندسة شبكات في ما يخص مادة التجارة الالكترونية؟

للإجابة على مثل هذه الأسئلة, والتي تخص المستقبل, نحتاج لتطبيق التحليلات التنبؤية. هنا يقوم الأستاذ باستعمال خوارزميات تعلم الآلة. ويختار منها خوارزميات تحت الإشراف (Supervised) من نوعية الانحدار (Regression). أما اختيار خوارزميات تعلم الآلة تحت الإشراف فيعود لامتلاكه لبيانات الخمس سنوات الفارطة مع الأعداد في مادة التجارة الإلكترونية التي نرغب في التنبؤ بها. أما اختيار خوارزميات الانحدار فيعود أن الأعداد هي ذات قيمة متصلة تتراوح بين 0 و 20. يقوم المدرس بتجربة عدد من الخوارزميات ويقارن بينها من

خلال مختلف المقاييس. ثم يختار الأفضل ويستعمله. ثم يقارن النتائج المتوقعة للطالب دون دراسة سلامة الشبكات ومع دراسة سلامة الشبكات. الرسوم البيانية تبين أن معدل الفرق في الأعداد هو في صالح دراسة سلامة الشبكات بفارق ثلاث نقاط في الأعداد.

في الختام يطرح المدرس سؤال: ماهي المواد التي يرجى إضافتها كمواد مشتركة من أجل تحسين الأعداد في مادة التجارة الإلكترونية؟ هنا قد يبدو الجواب بديهي وهو: يجب إضافة مادة سلامة الشبكات كمادة مشتركة !

لكن المدرس قرر أن يقوم بتحليلات الوصفة من أجل الإجابة على هذا السؤال. هنا تبين مختلف الرسوم البيانية أنه إضافة إلى مادة سلامة الشبكات فيجب إضافة مادة المبادئ الأساسية للشبكات أيضا. حيث تبين التحليلات أن الوصفة المثلى للنجاح في مادة سلامة الشبكات (وبالتالي النجاح في مادة التجارة الإلكترونية) يتطلب فهما جيدا لمادة المبادئ الأساسية للشبكات والتي هي أيضا من بين المواد الخاصة فقط بطلبة هندسة الشبكات.

هذا المثال العملي يبين:

- ماهية أنواع تحليلات البيانات
- الهدف من كل نوع منها
- الترابط بينها
- التمشي الأمثل لاستخدامها
- ضرورة طرح الأسئلة المناسبة قبل كل نوع من التحاليل

- أهمية الإلمام بمفاهيم وتقنيات تحليلات البيانات
- الأثر البليغ والقيمة المضافة لتطبيق تحليلات التعلم

3. جمع البيانات التعليمية

3. 1 المقدمة

هذا الجزء سيكون مخصص للبيانات وكيفية جمعها وذلك لاستخدامها من أجل التشخيص والتنبؤ. إذ أن جودة هذه البيانات ستحدد نوعية النتائج المتحصل عليها. سنبدأ بدراسة جمع البيانات في المساقات الهائلة المفتوحة على الإنترنت التي تسمى بالموك (MOOC : Massive Open Online Course) ثم بدراسة جمع البيانات في بيئة التعلم المعززة بالتكنولوجيا (TELE :Technology Enhanced Learning Environment) واخيرا بدراسة جمع بيانات القنوات المتعددة (MC : Multichannel). علما وأنه يجب أن يتم جمع البيانات مع احترام بعض الجوانب القانونية مثل حماية وأمن البيانات الشخصية (Ethics and Privacy) ولهذا السبب تم وضع مجموعة كاملة من الأخلاقيات لضمان جمعها بطريقة مهنية .

3. 2 جمع البيانات

يعد جمع البيانات خطوة أساسية لأنها تتضمن استعادة جميع البيانات الرقمية التي تمثل نشاط المستخدم من أجل إجراء عملية تحليل والحصول على انعكاس لتقدم مواقف التعلم. وبالتالي ، فإن جمع البيانات يتعلق بملاحظة المتعلم وتتبع تفاعلاته بوساطة أدوات التعلم المختلفة ، وتخزين الآثار التي تم جمعها واستيراد الآثار في أدوات التحليل.

تأتي هذه البيانات من عدة مصادر غير متجانسة مثل بيئات التعلم الرقمية القائمة على أنظمة إدارة التعلم مودل (Moodle) و بيئات الموك (Moocs) وبيئات التعلم المعززة بالتكنولوجيا (TELE : Technology Enhancement Learning Environment). لا ننسى أيضا أنظمة التعلم داخل المؤسسات التعليمية الأخرى التي يمكن أن تولد بيانات مفيدة لقياس أداء المتعلم مثل: المحاكاة والآلات الحاسبة.

فيما يلي ، سنشرح ماذا وكيف يتم جمع هذه البيانات على مستوى كل مصدر، والهدف هو الحصول على عينة ذات مغزى وجودة من البيانات للتحليل.

3. 3 جمع البيانات في المساقات الضخمة المفتوحة عبر الإنترنت

سنناقش في هذا المقطع كيفية جمع البيانات في بيئة الموك (MOOC : Massive Open Online Course)والذي يمكن ترجمته بالمساقات الضخمة المفتوحة عبر الإنترنت. افترض أنك مسؤول على الدورة التدريبية في الموك ولديك حق الوصول إلى جميع البرامج وجميع البيانات الموجودة ويمكنك كتابة البرامج التي تحتاجها لجمع البيانات الناتجة عن الطلاب على الموك. فما هي البيانات التي يمكن جمعها عن المتعلمين؟ هناك بيانات مشتركة عند الحديث عن الMOOC مثل الطابع الزمني لكل حدث أو نشاط (Timestamp) ، معرف المتعلم ، (Learner-ID) ، معرف الجلسة (Session-ID) والعنوان الخاص بالجهاز(IP).

يمكن للمتعلم تسجيل دخوله إلى الموك عدة مرات في يوم واحد أو على مدار الوقت لذلك فإن تسجيل معرف الجلسة وعنوان IP لمعرفة من أين يقوم المستخدم بتسجيل الدخول مهم. الغرض من هذا ، جمع البيانات لجميع المستخدمين عبر الموك لتحديد أي مستخدم أي جلسة ومن أين أتى والطابع الزمني للحصول على البيانات الدقيقة مثل وقت وقوع الحدث.

اعتمادًا على منصة الموك ، يمكن القيام بأنشطة مختلفة في منتدى مناقشة مثل: التعليق، الرد، التصويت، التصويت لصالح التعليق أو الإعجاب بالتعليق. يمكن ايضا حذف موضوع، إلغاء متابعة سلسلة رسائل، متابعة سلسلة رسائل أو تحديثها أو تفويتها.

هنا قدمنا البيانات الأساسية التي يمكن جمعها والتي توجد في ملفات السجل لأي بيئة تعليمية، ولا سيما الموك. بالإضافة إلى ذلك ، لدينا إمكانية جمع البيانات البيداغوجية التي تم إنشاؤها بواسطة المتعلم أثناء الأنشطة التي قام بها في الموك (آثار الأنشطة). في هذا الإطار يمكننا معرفة ما إذا كان الطالب قد قام بزيارة الموارد التعليمية الخاصة بموضوع ما وقراءتها أم لا؟ إذا كان الطابع الزمني مهم فيمكنك الحصول مثلا على مدة نشاط الطالب لمقارنتها بالحد الأدنى للقيام بهذا النشاط. في المنتدى ، يمكن للمتعلم البحث عن شيء ما أو مستخدم معين لذلك ، يجب تسجيل هذا النوع من الأنشطة في الموك. بصرف النظر عن هذا ، يتنقل الطالب في الموك من محتوى إلى محتوا آخر. في هذه الحالة، يمكن تسجيل سلوك التنقل لدى هذا الطالب.

نظرًا لأننا في بيئة تعليمية من صنف الMOOCs، فإننا نواجه عمومًا ثلاثة أشياء: العدد الكبير من الموارد والعدد الكبير من الكبسولات الرقمية وتسلسلات الفيديو ومنتديات المناقشة. لذلك ، من الضروري جمع تلك

البيانات لتقييم تفاعل وسلوك الطلاب. بمجرد جمع البيانات ، سنحتاج إلى تصفيتها حتى نتمكن من تفسيرها. للقيام بذلك ، يجب على المطورين كتابة البرامج المناسبة التي ستعالج هذه البيانات وتنتج مخرجات قابلة للتفسير.

في الختام ، سنقول أن البيانات التي سيتم جمعها في بيئة التعلم من صنف الMoocs ، هي معرف المتعلم ، ومعرف الجلسة ، و IP الخاص بالجهاز ، وجميع بيانات الأنشطة التي قام بها المتعلم في فترة زمنية (يجب أن يحتوي كل إجراء على طابع زمني، معرف المستخدم، ومعرّف الجلسة ، أيضًا ما هو اسم الإجراء : فيديو أم بحث أم أنه تمرير، هل هو إنشاء منتدى مناقشة، هل هو حذف منتدى المناقشة وسياق الإجراء). ستتم معالجة هذه البيانات بواسطة برامج مشفرة بأي لغة ؛ والغرض من ذلك هو تحويل هذه البيانات الأولية إلى معلومات قابلة للتفسير يمكننا استخدامها مع أدوات أخرى مثل جدول بيانات Excel أو خوارزميات تعلم الآلة أو غيرها من تقنيات تحليلات التعلم.

3. 4 جمع البيانات في بيئة التعلم المعززة بالتكنولوجيا

في هذا المقطع سوف نستعرض البيانات التي يجب جمعها في بيئة التعلم المعززة بالتكنولوجيا : (TELE Technology Enhancement Learning Environment) وهي بيئة تعليمية مختصة لتعليم مجال معين)على سبيل المثال ، بيئة محاكاة(. تحتوي البيئة مثل العديد من بيئات التعلم الأخرى على هدف تعليمي : المتعلم مطالب بالقيام بمهام معينة أو يتعين عليه حل مشاكل معقدة حتى يتعلم بعض المهارات أو عليه الإجابة على مجموعة من الأسئلة. في كل هذه الحالات ، تتوفر لديه موارد مثل مواد القراءة والأدوات التعليمية التي يمكنه استخدامها للوصول إلى الأهداف المذكورة. سنناقش واحدة من بيئة التعلم المعززة بالتكنولوجيا (TELE)

وتسمى بالميتل (MEttLE [26]: Modelling-based estimation learning environment). وهي بيئة تعليمية تم إنشاؤها لتعليم مهارات تقدير الكم من خلال الأشغال التطبيقية الهندسية لطلاب الهندسة الميكانيكية والهندسة الكهربائية في السنوات الثانية والثالثة بقسم تكنولوجيا التعليم التابع للمعهد الهندي للتكنولوجيا بومباي (IIT Bombay) .

عندما يقوم الطالب بتسجيل الدخول إلى MEttLE، سنحصل على التعليمات الاولية الخاصة به (تاريخ وساعة الدخول ومعرف الطالب). بعد المصادقة (Authentication)، يمكن للطالب البدء في حل مشكلة معقدة والتي تقسم إلى مشاكل فرعية أصغر. لذلك، يمكن للطالب إنشاء حل للمشكلات الفرعية والجمع بين الحلول لحل المشكلة الرئيسية. هنا يعطي MEttLE[27] المشاكل الفرعية للتقدير الهندسي والتي تكون ضرورية لكسب مهارة التقدير الهندسي .عند تحديد موضوع فرعي ، سيتم تزويد الطالب بمجموعة من الأسئلة للإجابة عليها وتقييم ما إذا كان قد حل المشكلة بالفعل أم لا.

هناك أدوات أخرى يستعملها الطالب اثناء التعلم مثل المحاكي (Simulator)، حيث يمكن للطالب تشغيل فيديو المحاكاة ومحاولة فهم ما يجري. على غرار المحاكاة ، هناك بعض الأدوات الأخرى مثل الآلة الحاسبة (Calculator) أو لوحة الخربشة (Scribblepad). باستخدام الآلة الحاسبة ، يمكن للمتعلم تغيير قيمة المتغيرات ويمكنه المحاكاة مرة أخرى لمعرفة التأثيرات بعد تحديث المتغيرات.

26 MEttLE: a modeling-based learning environment for undergraduate engineering estimation problem solving.

27 MEttLE: a modeling-based learning environment for undergraduate engineering estimation problem solving

بما أن لدينا الآن فكرة كافية عن طريقة عمل ,MEttLE من الواضح أن بحوزتنا إذا كمية كبيرة من البيانات التي تم إنشاؤها بواسطة المتعلمين من خلال استخدامهم لتلك البيئة. فما هي ماهية ونوعية تلك البيانات؟

مثلما ناقشنا في بيانات بيئة الموك (MOOC)، هناك معلمات (Parameters) شائعة مثل الطابع الزمني لكل حدث أو إجراء، معرف المتعلم، ومعرف الجلسة. بالإضافة لهذه البيانات تختص بيئات الTELE بوجود أنواع أخرى من البيانات المتاحة. من ضمنها تلك التي يمكن جمعها حول استخدام الأدوات)مثل المحاكي والحاسبة) من قبل المتعلم. كما يمكن جمع السلوكيات التي لوحظت أثناء استخدام مثل هذه الأدوات على غرار تغيير الرسم البياني أو تغيير قيمة المتغيرات. نفس الشيء وبما أن الطالب يمكنه التنقل إلى صفحات متعددة، فيجب تتبع آثار هذا التنقل وكذلك جمع الإجابات عن الأسئلة المطروحة في كل مشكلة فرعية. كل هذه البيانات التي تخص الطالب يجب أن يتم التقاطها وجمعها وتخزينها ومعالجتها من أجل تحليلها والاستفادة منها.

3. 5 جمع بيانات القنوات المتعددة

لقد ناقشنا, في الأقسام السابقة، كيفية جمع البيانات من بيئة الفصل الدراسي (Classroom Environment)، و من المساقات الضخمة المفتوحة عبر الإنترنت MOOC)) و من بيئات التعلم المعززة بالتكنولوجيا .(TELE)

إضافة إلى البيانات التقليدية المتوفرة في مختلف بيئات التعلم, فإنه من الممكن, وحسب الهدف المنشود, التوسع في جمع البيانات. فقد ذكرنا في الباب الثاني من هذه الدراسة (تحليلات البيانات) أنه بالإمكان بالإضافة للبيانات الداخلية استخدام بيانات خارجية (حالة الطقس, اضطرابات مواقيت وسائل النقل, الأحداث الهامة الشعبية, ...) من أجل تحليلات أكثر شمولا قد تهم مثلا دراسة حالات تغيب الطلاب. لن نخوض أكثر في هذا المبحث بما أنه قد أخذ القسط الكافي من الشرح كما ذكرنا سابقا. لكن من المهم الإشارة إلى تعدد التجارب التي تسعى إلى جمع نوع جديد من البيانات باستعمال معدات خاصة وذلك في إطار تحليلات التعلم. هنا نتحدث عن جمع البيانات من قنوات متعددة (Multichannel).

يمكن مثلا جمع البيانات باستخدام المستشعرات ((Sensors. حيث تسمح هذه الأخيرة مثلا بجمع بيانات حول نظرة العين وتعبيرات الوجه وموجات الدماغ وغيرها من السمات. بالنسبة لبيانات نظرة العين مثلا، يمكن جمعها باستخدام متتبِّعات العين (Trackers)، بما ان هناك أجهزة تعقب لتحركات العين ويمكن توصيلها لجهاز الكمبيوتر بغرض الحصول على نظرة دقيقة للطلاب. كما يمكن أيضا استخدام كاميرا الويب الخاصة بالكمبيوتر لجمع بيانات نظرة العين.

إذا يمكن جمع البيانات من قنوات متعددة (Multichannel) مما يجعل معالجتها تحديًا لأننا بحاجة إلى محاذاة البيانات من كل قناة بناءً على الطابع الزمني. على سبيل المثال هناك أجهزة تتبع العيون يمكنها تسجيل نظرة عين الطلاب بتردد 1500 هرتز. مع العلم أنه إذا تم جمع نظرة العين بسرعة 60 هرتز ، فهذا يعني أن متتبع العين يجمع 60 إشارة في الثانية, فماذا عن كمية البيانات التي يمكن التقاطها بتردد 1500 هرتز. وماذا إذا تعددت القنوات. هنا نرجع إلى أهمية الإستفادة من تكنولوجيا البيانات الضخمة وتكنولوجيا الحوسبة السحابية.

3. 6 الأخلاقيات وخصوصية البيانات

لقد رأينا أنه يمكننا جمع الكثير من البيانات من بيئات التعلم والفصول الدراسية وعبر الإنترنت وبيئات التعلم المعززة بالتكنولوجيا وأيضًا من مصادر أخرى مثل بيانات نظرة العين وتعبيرات الوجه. ولكن لقياس وجمع وتحليل بيانات المتعلم ، يجب أن نحصل على موافقة المتعلم. إذا قمنا بالإبلاغ عن بيانات البحث إلى المعلم، فيمكننا شرح ما حدث للفصل وما يمر به الطالب ويمكن حتى تحديد اسم الطالب ومعرِّف الطالب وهذا لا يمثل إشكالا. ولكن إذا أردنا نشر البيانات للعموم، فما نوع البيانات التي يمكننا نشرها؟ وما هي الخطوات الأساسية التي يجب اتخاذها لحماية خصوصية البيانات والمتوافقة مع الجوانب الأخلاقية للنشر؟

عند جمع البيانات من متعلم ، فهنالك حاجة إلى الحصول على موافقته على إعطاء البيانات ومشاركتها. أما إذا كان المشارك قاصرًا فيجب الحصول على نموذج الموافقة من وليِّهِ .ثم عند نشر البيانات، يتعين إخفاء الهوية (الأسم, عنوان (... ,IP., هذا البروتوكول مهم جدًا عند نشر البيانات للعموم.

إزالتها أو حذفها.

تتعلق الأخلاقيات وخصوصية البيانات أيضًا بالتخزين. عند تخزين البيانات في السحابة أو في أقراص أو في جهاز حاسوب، يجب تأمين تلك البيانات .المقصود بالتأمين هنا هو ضمان عدم الوصول إلى البيانات لأي شخص غير مصرح له بالاطلاع عليها, وضمان عدم نسخها أو نقلها.

نذكر أنه في تحليلات التعلم، يتم جمع بيانات المتعلم بغرض تطوير نموذج يمكن من مساعدته على تحقيق النجاح في البيئة التي يعمل فيها. إذا كانت بيئة التعلم المعززة بالتكنولوجيا ، فيتم جمع بيانات الطالب مثل سلوكه في حل المشكلات والوقت الذي يقضيه في مشاهدة مقاطع الفيديو, ثم يتم إنشاء نموذج يسمح بتقديم المساعدة, كالملاحظات التصحيحية للمتعلم لمساعدته على تحقيق النجاح. في هذا الصدد هناك حاجة لسؤال الطالب في بعض الأحيان وفي أحيان أخرى يتم جمع البيانات بطريقة غير مباشرة من خلال تتبع وجمع آثار استعماله للبيئة التعليمية .هنا يجب دائما الحرص على إبلاغ المتعلم بكل ما يدور حوله. إذا يجب أن يكون لدى المستخدمين معرفة بالبيانات التي يتم جمعها ويجب الحصول على موافقته.

في نفس السياق، من أخلاقيات جمع البيانات السماح للمستخدم برفض الانخراط, أو إذا قام بالانخراط سابقا, السماح له بالانسحاب من عملية جمع البيانات. هنا يتعين على القائمين على هذه العملية حذف جميع البيانات التي جمعت عن المستخدم. والأهم من ذلك ، يجب ألا يعيق هذا الخيار المستخدم من استخدام البيئة التعليمية. يجرنا هذا الحديث إلى طرح السؤال التالي: ما هي حقوق المتعلم على بياناته في تحليلات التعلم؟

لقد رأينا أنه يجب إبلاغ المتعلم بالبيانات التي سيتم جمعها ، ولكن ما هي حقوق المتعلمين على بياناتهم في تحليلات التعلم. يحق للمتعلم الانسحاب من العملية في أي وقت أثناء الدراسة ويمكنه أن يطلب محو البيانات بالكامل أثناء الدراسة. وأيضًا يجب أن يكون لديهم إمكانية الوصول إلى بياناتهم إذا كانوا يريدون الاطلاع عليها. والأهم من ذلك أنه يجب السماح للمتعلمين بأن يكونوا جزءًا من الخوارزمية الخاصة بهم : على سبيل المثال ، إذا أعطيت الملاحظات، فلا يجب أن تفرض الملاحظات على المتعلم بل يجب أن يكون لديه خيار قبول أو رفض الملاحظات.

للإشراف على هذا الأمر وفي إطار الحوكمة الرشيدة للمؤسسات التعليمية فمن المهم أن تكون هنالك لجنة تعنى بالأخلاقيات في مجال جمع البيانات. هذه اللجنة يجب أن تقوم بدور استشاري ودور تحكيمي ودور توعوي ودور تشريعي:

- الدور الإستشاري الهدف منه هو توجيه ونصح وتفسير الحقوق والواجبات للمستخدمين وللراغبين في جمع وتحليل البيانات. هذا سيسمح باستباق المشاكل والنزاعات وتجنب وقوعها.
- الدور التحكيمي الهدف منه هو فض المشاكل والنزاعات حول جمع وتحليل البيانات إذا وقعت. كما عليها إيجاد الحلول التحكيمية والحرص على تطبيقها.
- الدور التوعوي الهدف منه نشر الثقافة المتعلقة بأخلاقيات جمع وتحليل البيانات والمخاطر المتعلقة بها. كما تسهر أيضا على نشر ثقافة سلامة البيانات وحماية الخصوصية.
- الدور التشريعي يهم أولا جمع وتبويب ونشر النصوص القانونية ذات الصلة. كما يهم أيضا إعداد ميثاق سلوك خاص بالمؤسسة التعليمية متوافق مع القوانين ذات العلاقة و متلائم مع خصوصيات المؤسسة.

4. أدوات تحليلات البيانات واستخداماتها

4. 1 لوحة القيادة

تعتبر لوحة القيادة (Dashboard) من أهم أدوات تحليلات البيانات عموما وتحليلات التعلم خصوصا. ويتم استخدامها أساسا في إطار التحليلات الوصفية التي يقال عنها تحليلات لوحة القيادة (Dashboard

Analytics). تمثل هاته اللوحة البيانات المعقدة ممثلة في رسومات سهلة الاستخدام مثل الرسوم البيانية أو المخططات الدائرية أو الرسوم البيانية التفاعلية أو بعض الرسوم البيانية المقارنة. يعني ان هاته اللوحة تقوم بجمع بيانات المتعلمين (في إطار تحليلات التعلم) أو المستخدمين أو الموظفين (في إطار تحليل البيانات عموما) وتمثيلهم في لوحة معلومات لتحليلها من أطراف أصحاب المصلحة واتخاذ الإجراءات والقرارات اللازمة بناءا على ذلك.

في تحليلات التعلم ، لوحة القيادة تمثل اداة مهمة للباحثين والمعنيين بالتحليلات بصفة عامة، لاستخراج الاستنتاجات والأفكار لتحديد ما حدث. لذلك ، هذه هي الخطوة الأولى عادة عندما تقوم بالبحث عن البيانات وتجميعها, حيث يقع عرضها في شكل رقمي أو في شكل رسومات. تُستخدم لوحة القيادة أيضًا لتبليغ الأفكار أو النتائج التي استخرجها محلل البيانات إلى أصحاب المصلحة الآخرين مثل المدرسين أو المتعلمين.

في التحليلات الوصفية أو تحليلات لوحة المعلومات، يجب أن تكون البيانات الممثلة دقيقة للغاية. لتحقيق هذه الجودة، يجب الإجابة عن الأسئلة الرئيسية التالية :

- ما هي البيانات التي يجب إظهارها؟
- لماذا نحتاج إلى إظهار هذه البيانات؟
- ما هي البيانات التي نريد الاطلاع ؟
- من هم المستخدمون المعنيون بتلك البيانات؟
- ما هي تقنية التصور (Visualization) التي سنستخدمها لإظهار البيانات؟

لنفترض أن لدينا بيئة تعليمية مثل MEttLE[28] التي قدمناها سابقا أو مساق مفتوح عالي الإستقطاب MOOC أو مقاطع فيديو YouTube, هنالك الكثير من البيانات التي يمكن جمعها من هذه البيئات. علينا فحص وتحليل هذه البيانات, وتمثل لوحة القيادة أداة مهمة لإدراك هذه الغاية. فما هي المؤشرات التعليمية التي من الممكن عرضها؟ يمكننا ذكر معيار الأداء في الاختبارات في البيئات التي ذكرناها أو مستوى الأداء في نهاية الفصل الدراسي. كما يمكن عرض الوقت المقضى في كل مهمة أو نشاط. كما يمكن أن نعرض البيانات التي تسمح بتحليل مدى استخدام الموارد التعليمية. هذا سيسمح بمتابعة مدى استخدام الأدوات كالمحاكي أو الآلة الحاسبة. كما قد يسمح بدراسة استخدام محتوى ما للإجابة عن أسئلة على غرار: ما هو الفيديو التعليمي الأكثر مشاهدة؟ ما هي مدة مشاهدة الفيديو؟ كم عدد مرات مشاهدة الفيديو؟ ما هي سرعة مشاهدة الفيديو؟ كل هذا يمكن أن يمثل مجموعة هامة من البيانات لتمثيلها في تحليلات لوحة المعلومات.

من البيانات التعليمية الأخرى التي يتم الحرص على متابعتها من خلال لوحة البيانات هي المحتويات التي تم إنشاؤها من قبل المتعلم. فعلى سبيل المثال يمكن متابعة عدد المشاركات التي ينشرها المتعلم في منتدى المناقشة, أو عدد المنتديات التي أنشأها المتعلم, أو عدد الأسئلة التي طرحها المتعلم في منتدى المناقشة. أيضا يمكن متابعة تسليم تقارير العمل والواجبات المستحقة ومدى احترام الآجال المحددة وعدد الواجبات التي سلمها الطالب.

لوحة البيانات عمليا يمكن اعتبارها نموذجا يسمح بإظهار البيانات المعقدة بطريقة تسهل الاستخدام. حيث تعتمد على عرض البيانات بطريقة مرئية اعتمادا على الصور أو الرسوم البيانية أو المخططات الخطية أو المخططات الشريطية وذلك من أجل المساعدة على فهم البيانات الخاصة ببيئة التعلم.

28 MEttLE: a modeling-based learning environment for undergraduate engineering estimation problem solving

من المهم للحصول على رسوم بيانية ذات مغزى، اختيار الأسئلة المناسبة التي ستظهر نتائج قابلة للتفسير.على سبيل المثال يمكن أن نسأل: ما هي نسبة نجاح الطلاب في الفصل على مدار 3 سنوات؟ يمكن ببساطة عرض رسم بياني يوضح نسبة النجاح على مدار السنوات الثلاث الماضية . يمكن ايضا الاستشهاد بالأسئلة الأخرى التي أثارها الباحثون وأثبتت أهميتها إذا أردنا في أي وقت عرض نتائجها في رسوم البيانية :

- ما هي نسبة الحضور في الفصل في فترة معينة من العمل؟
- هل الحضور يتناقص أم أن الحضور ثابت على مدار هاته الفترة؟ أم أن هناك تباينًا كبيرًا في الحضور؟

الوقت المقضي من قبل المتعلم في دراسة الموارد التعليمية هو أيضًا عامل رئيسي مهم لتمثيله في رسم بياني. فعلى سبيل المثال ، إذا كان لدينا خمسة موارد تعليمية، فيمكن تمثيل معدل الوقت الذي يقضيه جميع الطلاب في كل مورد مما يعطي مزيدا من الأفكار للباحث.

يمكن كذلك تحديد معدل الإشراف على الطلاب من طرف أعضاء هيئة التدريس. ثم يمكن اذا مقارنة مدى تأثير هذا المعدل على النتائج. هل العائد أو الأداء أفضل أم لا؟ فعلى سبيل المثال ، نسبة الطلاب إلى أعضاء هيئة التدريس في حالة وجود فصلين، إذا كان الفصل ب يحتوي على نسبة أفضل من أعضاء هيئة التدريس بالفصل أ، هل يكون أداء المتعلم أفضل هناك؟ يمكننا الوصول إلى هذا الاستنتاج بعد عمل ارتباط بين الوسائل المتعلقة بالفصلين وتمثيل نتائج التحليل بواسطة رسم البياني.

بصفة عامة ، إذا قررت سؤالاً ما، فيجب أن تفكر أولاً من أين يمكن جمع البيانات للإجابة على هذا السؤال. يتعين عليك بعد ذلك تنظيف البيانات, ثم تحليلها. ويمكن مراجعة الفصل الثاني من هذه الدراسة, تحليل البيانات, حول مختلف هذه النقاط.

4. 2 العرض المرئي للبيانات في لوحات القيادة

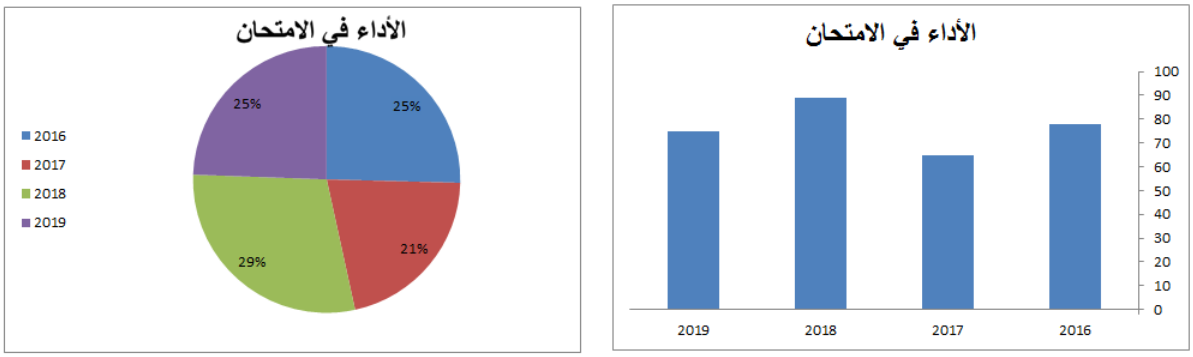
سنتحدث هنا عن أنواع العرض المرئي للبيانات ومدى ملائمتها عند البحث عن جواب لسؤال بحث معين (RQ :Research Question). إذ أن النتائج التي تنتج عن التحليلات يتم عرضها عادة من خلال لوحة القيادة عبر البيانات المرئية. هنا يمكن طرح العديد من الأسئلة على غرار: ما هو نوع الرسم البياني الذي يفضله المستخدمون بشكل أفضل؟ ما هو الرسم البياني الذي يحتوي على عبء معرفي أقل؟ أي مخطط يعطي مزيدًا من الأفكار حول البيانات؟

بحسب الدراسات العلمية التي أجريت، يفضل بشكل عام استخدام المخططات الشريطية مقارنةً بالمخططات الدائرية. يسمح عادة المخطط الشريطي الذي يحتوي على القيم أو المخططات التفاعلية بالحصول على القيمة المضافة. يُفضل هذا النوع من الرسم البياني للمستخدمين مقارنةً بالمخططات الشريطية العادية. أبحاث أخرى تقول أن الرسوم البيانية ثلاثية الأبعاد لا تقدم إضافة كبيرة مقارنة بالرسومات البيانية العادية. لتوضيح ما شرحناه، سنقدم الأمثلة التالية:

- في الرسم البياني التالي ، تُظهرالبيانات ببساطة النسبة المئوية للنجاح أو الأداء في الاختبار للسنوات الأربع الماضية من 2016 إلى 2019. يمكنك أن نرى بوضوح في الرسم البياني

الشريطي أن أداء الطلاب في عام 2016 بلغ 78٪ ، وفي 2017 65٪ ، وفي 2018 كان أكثر مثل 89٪ وفي عام 2019 كانت النسبة أقل قليلاً. في الجانب الآخر ، عندما تنظر إلى هذه البيانات في مخطط دائري ، لا يمكننا تحديد ما تمثله هذه البيانات لأن المخطط الدائري يمثل النسبة المئوية للقيمة ككل، ولا يمثل القيم على مدار الفترة الزمنية. لذلك ، عندما تقوم بتمثيل البيانات على مدى فترة زمنية ، لا تستخدم المخطط الدائري ،بل من الأفضل استخدام المخطط الشريطي.

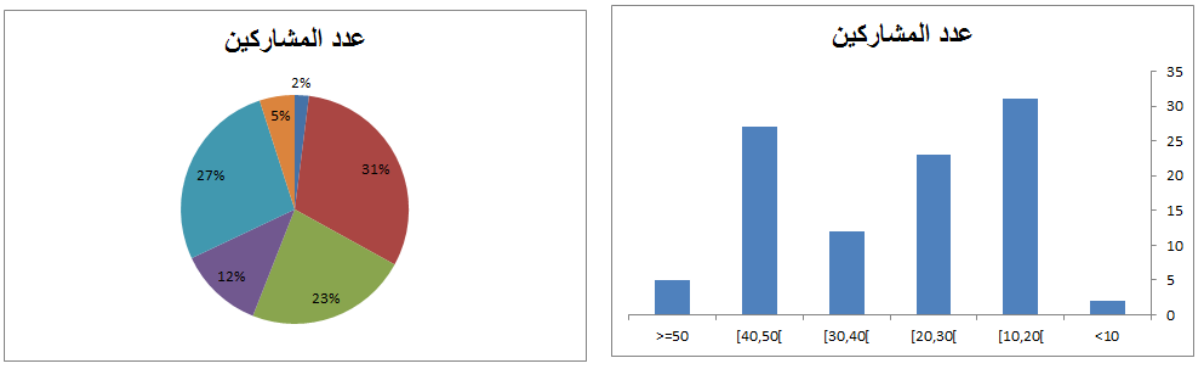
رسم البياني الشريطي والمخطط الدائري



أهمية اختيار الرسم البياني

- مثال آخر، حيث يمثل الرسم البياني لعدد المشاركين في مساق تعليمي مفتوح عبر الإنترنت في استطلاع أجري عليهم. لذلك، إذا نظرت إلى عدد المشاركين الذين شاركوا في الاستطلاع أو الاستبيان، ستجد مشاركين فقط تقل أعمارهم عن 10 سنوات، بالنسبة للعمر بين 10 و 20 عامًا، ستجد 31 مشاركًا وبالمثل بالنسبة للفئة العمرية الأخرى. في هذا المخطط الشريطي المتغير الفئوي يعطي بيانات جيدة جدًا كما يعطي الرسم البياني الدائري نفس البيانات ذات المعنى. هنا إذا الاختيار بين نوعي المخططات لا يصنع فرقا جوهريا. هنا ما نود الإشارة إليه هو أهمية اختيار الألوان المناسبة. فمثلا هنا اللون الأحمر يعطي جاذبية أكثر مقارنة بالألوان

الأخرى, وهو يشير إلى عدد المشاركين في سن 10 إلى 20. وهذا من شأنه تبسيط عملية قراءة البيانات.



مثال لتوضيح أهمية اختيار الرسم البياني وأهمية استخدام الألوان

لقد رأينا بشكل عام مسألة استخدام الرسوم البيانية والألوان من أجل العرض المرئي للبيانات في لوحات القيادة . فما هي بعض استخداماتها بالنسبة للمتعلمين والمعلمين والباحثين؟

بالنسبة للمتعلمين لوحات القيادة تسمح بفهم وقياس أدائهم التعليمي. يمكن عرض رسم بياني شريطي لأداء المتعلمين خلال فترة زمنية, إذ أن نسبة التقدم مثلا تمثل عادة محفزا هاما الاستمرار والمواصلة .

بالنسبة للمعلمين أيضا لوحة القيادة مهمة جدًا، فهي تسمح مثلا بإلقاء نظرة مجردة للأداء والتفاعلات في الفصل. لنأخذ مثال عملي لاستخدام لوحة التحكم من قبل المعلمين. في هذا المثال يعمل الطلاب في بيئة تعليمية معينة، ويريد المعلم معرفة عدد الطلاب الذين تمكنوا من حل المشكلة 1، وعدد الطلاب الذين قاموا بحل المشكلة 2 ، وكم عدد الطلاب الذين تمكنوا من حل المشكلة 3. من خلال نظرة عامة على الرسوم البيانية, يمكن أن يجزم مثلا بأن المشكلة رقم 3 تشكل عائقا للمتعلمين. بالتالي يمكن له أن يبحث عن الأسباب ويعالج الأمر مما

يسمح بتحسين المساق التعليمي ومردود المتعلمين .أيضا يحتاج المعلمون إلى إلقاء نظرة للأداء والتفاعلات في الفصل بصفة فردية .البيانات الفردية لكل طالب تمكن المعلم من التدخل عند الحاجة مما يسمح بمساندة وتوجيه ودعم المتعلم الذي يشكو من صعوبات.

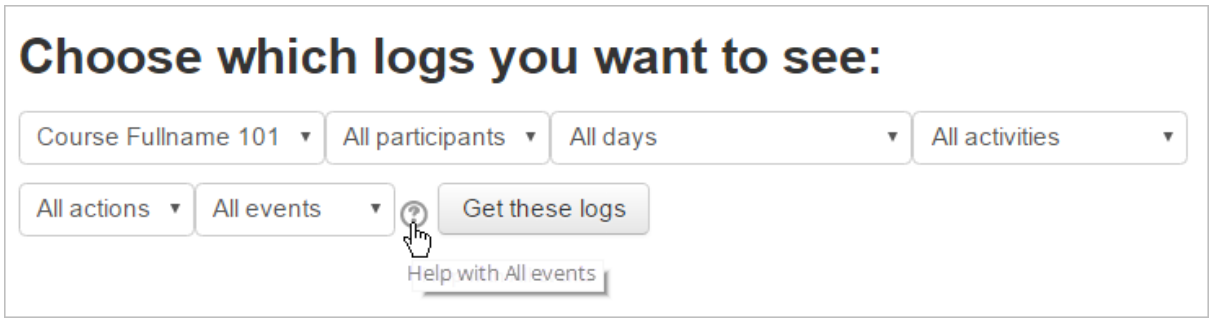
بالنسبة للباحثين، تعد لوحات القيادة هامة جدًا لأنه المكان الأنسب الذي يحاول فيه الباحث اكتشاف الأنماط التي تسمح بفهم ما يحدث في بيئة تعليمية معينة، أو في دورة تدريبية معينة. من خلال النظر في أنماط البيانات، والنظر في تصور البيانات، يمكن للباحثين استخراج العلاقات بين البيانات. نقطة الانطلاق تكون كالعادة سؤال بحثي. مثلا, السؤال قد يكون: هل هناك علاقة بين عدد أعضاء هيئة التدريس و الأداء؟ إذا ، هذا النوع من الأسئلة البحثية يأتي من خلال فحص ودراسة البيانات التي تقدمها لوحات القيادة .ثم بناءا على ذلك يقوم الباحثون بجمع الأفكار وتبليغها إلى مختلف المتدخلين في العملية التعليمية.

4. 3 لوحة القيادة في **MOODLE**

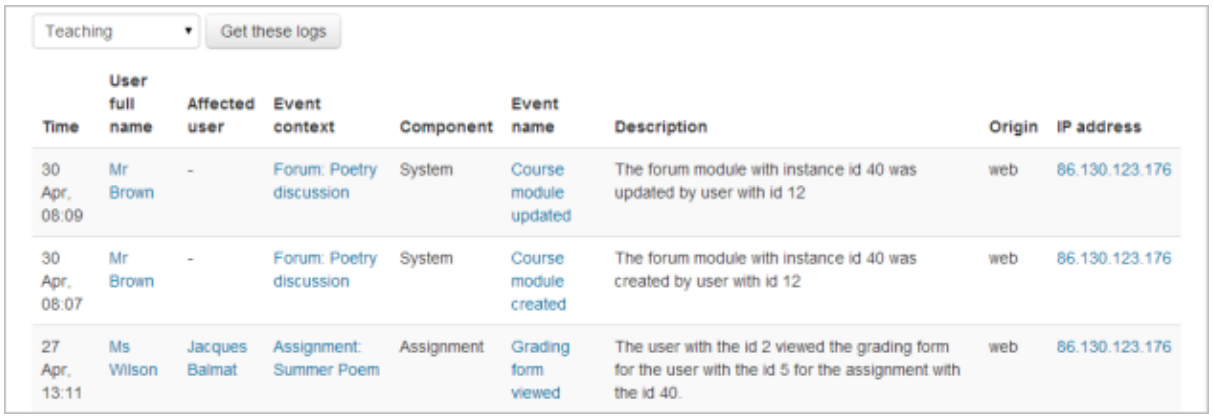
يعتبر MOODLE أحد أكثر منصات التعليم الإلكترونية انتشارا وذلك لعدة أسباب. من ضمن أسباب انتشاره أنه مفتوح المصدر, مجاني,مع إمكانية اضافة وحدات إضافية إليه حسب الحاجة. من ضمن الوحدات الإضافية نجد جملة من الأدوات الخاصة بتحليل البيانات والتي تسمح بإثراء لوحة القيادة. في ما يلي سنذكر بعضها:

- Logs[29]

تسمح هذه الأداة بالاطلاع على جملة من بيانات التفاعل مع المنصة التعليمية أو درس محدد. يمكن غربلة البيانات للاطلاع على تفاعل المدرسين، أو المتعلمين. يمكن استعماله من قبل المدرسين أو مديري الموقع أو صناع القرار. مخرجات هذه الأداة عبارة عن تقرير يمكن تصديره خارج المنصة.



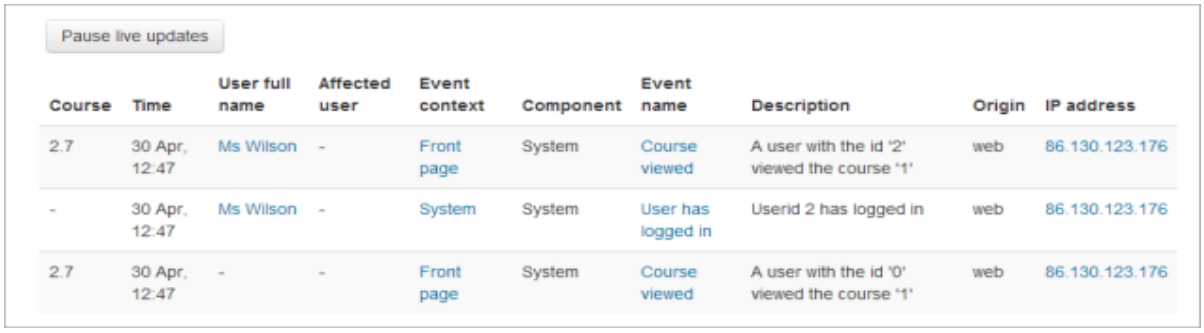
اختيار المعلمات الخاصة بالتقرير لأداة Logs



نبذة عن التقرير الذي تنتجه أداة Logs

كما تسمح الأداة أيضا بمتابعة مباشرة لبيانات الاستعمال.

29 https://docs.moodle.org/311/en/Logs



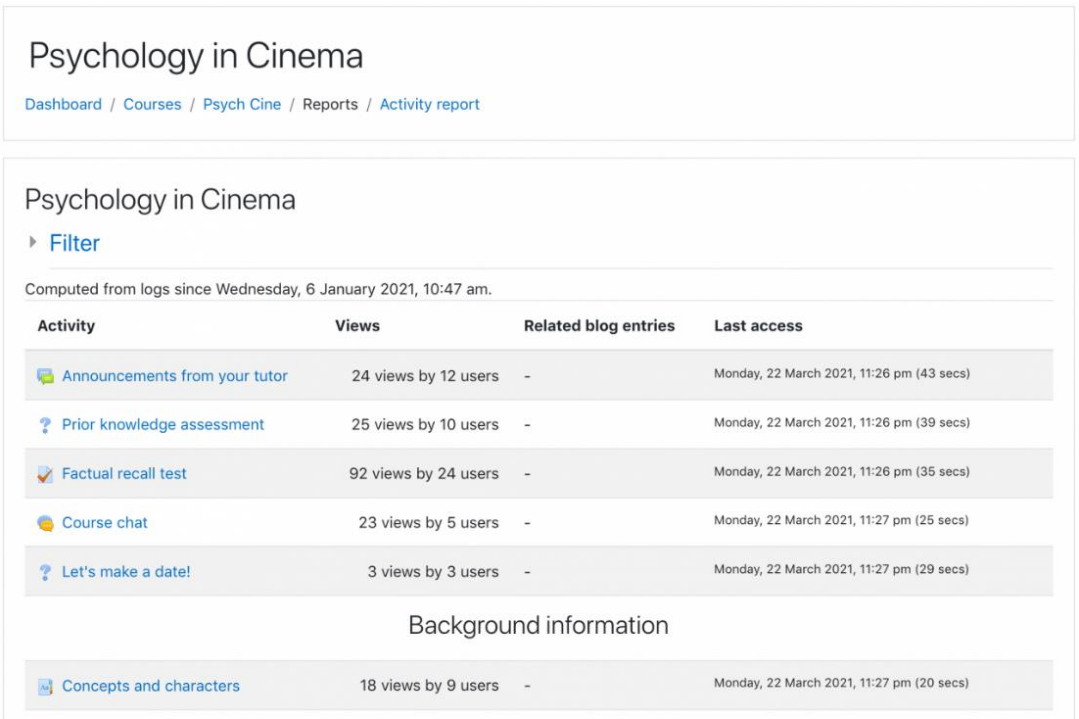
Pause live updates

Course	Time	User full name	Affected user	Event context	Component	Event name	Description	Origin	IP address
2.7	30 Apr, 12:47	Ms Wilson	-	Front page	System	Course viewed	A user with the id '2' viewed the course '1'	web	86.130.123.176
-	30 Apr, 12:47	Ms Wilson	-	System	System	User has logged in	Userid 2 has logged in	web	86.130.123.176
2.7	30 Apr, 12:47	-	-	Front page	System	Course viewed	A user with the id '0' viewed the course '1'	web	86.130.123.176

نبذة عن تقرير المتابعة المباشرة للبيانات لأداة Logs

- Activity report[30]

تسمح هذه الأداة بإحصاء عدد المشاهدات لكل محتوى أو نشاط وتقديمها للمشرفين والمدرسين. كما يمكن استعمال الأداة لتمكين المتعلم من الاطلاع على إحصائيات خاصة بمشاركته في الدرس.



Psychology in Cinema

Dashboard / Courses / Psych Cine / Reports / Activity report

Psychology in Cinema

Filter

Computed from logs since Wednesday, 6 January 2021, 10:47 am.

Activity	Views	Related blog entries	Last access
Announcements from your tutor	24 views by 12 users	-	Monday, 22 March 2021, 11:26 pm (43 secs)
Prior knowledge assessment	25 views by 10 users	-	Monday, 22 March 2021, 11:26 pm (39 secs)
Factual recall test	92 views by 24 users	-	Monday, 22 March 2021, 11:26 pm (35 secs)
Course chat	23 views by 5 users	-	Monday, 22 March 2021, 11:27 pm (25 secs)
Let's make a date!	3 views by 3 users	-	Monday, 22 March 2021, 11:27 pm (29 secs)
Background information			
Concepts and characters	18 views by 9 users	-	Monday, 22 March 2021, 11:27 pm (20 secs)

30 https://docs.moodle.org/311/en/Activity_report

نبذة عن تقرير أداة Activity report

- Course completion, [31]Activity completion[32]

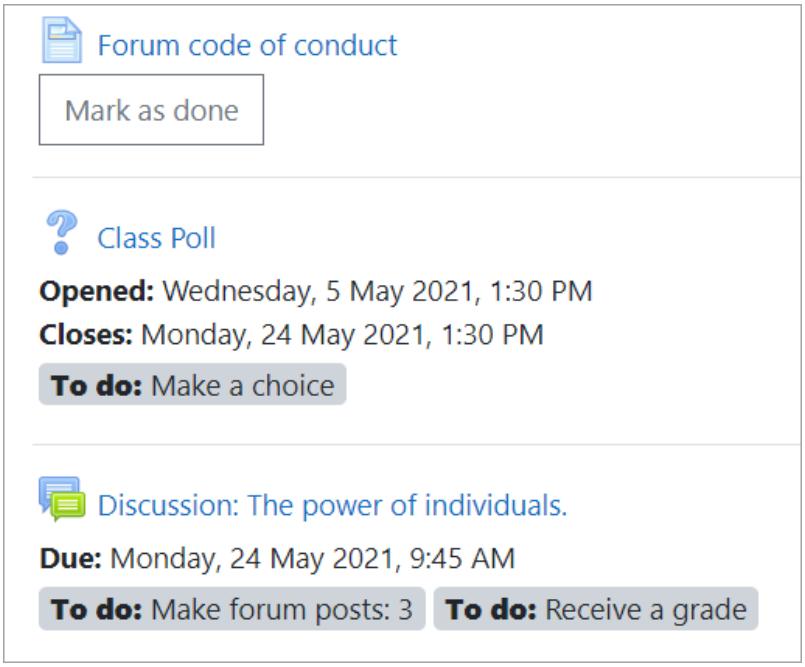
هما أداتان تتكاملان، فأداة Course completion تسمح للمدرس بمتابعة مدى تقدم المتعلم في استكمال الدرس. أما الأداة الثانية فتسمح للمتعلم بمتابعة مدى تقدمه في الدرس.

Criteria group		Dependencies	Activities		Grade	Approval	Course
Aggregation method		Any	Any		-	Any	All
Criteria		Course Completion 01	Lesson Activity Course ...	Quiz in Course completion 02	70.00 required	Teacher	Course complete
First name / Surname	Email address						
Test User1	testuser1@test.test	☑	☑	☑	☑	☐	☐
Test User2	testuser2@test.test	☑	☑	☐	☐	☐	☐
Test User3	testuser3@test.test	☑	☐	☐	☐	☐	☐
Test User4	testuser4@test.test	☐	☐	☐	☐	☐	☐

نبذة عن تقرير أداة Course completion

31 https://docs.moodle.org/311/en/Using_Course_completion

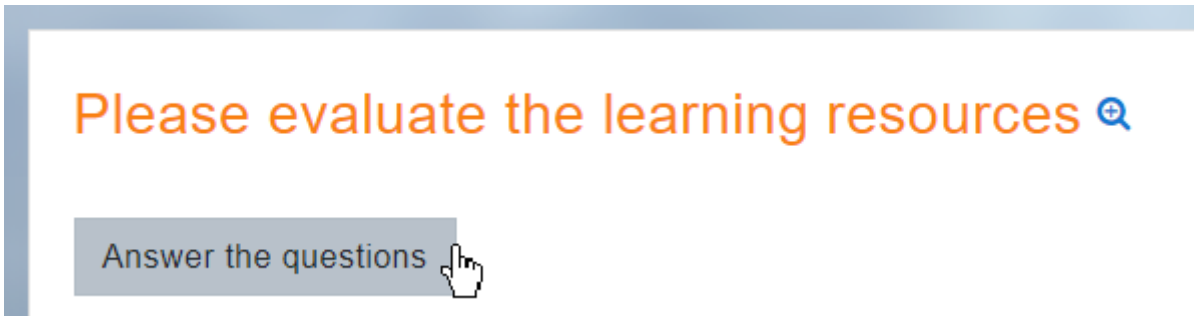
32 https://docs.moodle.org/311/en/Using_Activity_completion

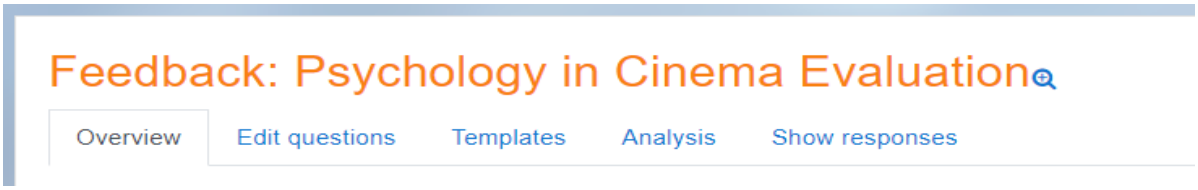


Activity completion نبذة عن تقرير أداة

- Feedback activity[33]

تسمح هذه الأداة للمدرسين باستبيان الطلبة حول الدرس المقدم كوسيلة للقيام بالمراجعات اللازمة.





Feedback activity واجهة صناعة استبيان لأداة

33 https://docs.moodle.org/311/en/Feedback_activity

واجهة الاطلاع على نتيجة الاستبيان لأداة Feedback activity

هذه بعض الأدوات المفيدة والمتعلقة بتحليلات التعلم والتي يمكن دمجها في بيئات التعلم MOODLE. هنالك العديد من الأدوات الأخرى المتوفرة والتي لم نذكرها في هذه الدراسة. إضافة لتلك التي هي قيد التطوير وقد تظهر قريبا لمزيد إثراء باقة أدوات تحليلات التعلم المتوفرة حاليا على نظام MOODLE.

4. 4 لوحة تحليلات يوتيوب

في هذا القسم ، سنناقش لوحة YouTube Analytics التي توفرها Google وهي لوحة تحكم التحليلات لمنشئي الفيديوهات على منصة YouTube كمثال متاح للجميع عن لوحات القيادة. تتيح لوحة القيادة للمطورين الإجابة عن العديد من الأسئلة: لماذا لم يتم جذب المزيد من المشاهدين لفيديو معين؟ ما هو سلوك المستخدمين عند مشاهدة مقاطع الفيديو؟

تمكن اذا هاته اللوحة المطورين من الحصول على أفكار تمكنهم من تطوير استراتيجياتهم من أجل زيادة المشاهدة لدى الجمهور. لتوضيح كل هذا نقترح المثال التالي لإحدى قنوات اليوتيوب .[34]أول سؤال يتبادر للذهن هو : ما هي البيانات التي يمكن جمعها من المشاهدين ؟ من ثم يمكننا استنتاج السؤال الموالي: هل هذه البيانات ستكون مفيدة لصاحب القناة؟

34 https://www.youtube.com/channel/UCtUBnuU8stYt1m6VniV8z5Q

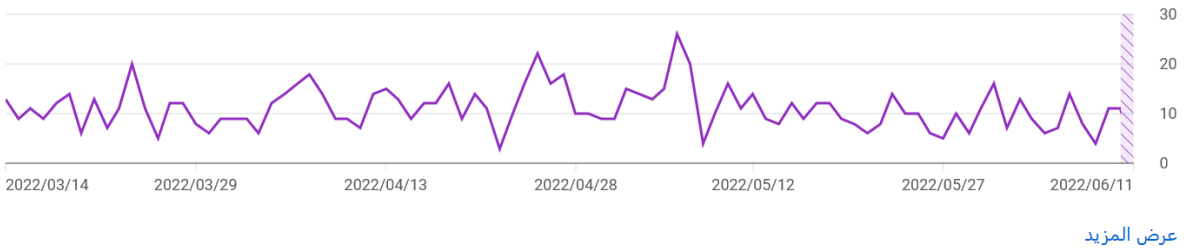
عادة يتم جمع ثلاث بيانات رئيسية : الاولى هي سلوك المشاهدة من خلال عدد المشاهدات للفيديو. الثانية هي البيانات المتاحة عن عن المشاهدين مثل العمر، الجنس وغيرها من البيانات. الثالثة هي البيانات التقنية عن المشاهدين كعنوان IP الخاص بالجهاز المستخدم عند مشاهدة الفيديوهات. مختلف هذه البيانات وغيرها متاحة من خلال لوحة القيادة. ستخبرك اللوحة عدد المشاهدين من البلد الذي تعيش فيه ، وعدد المشاهدين من البلدان الأخرى، ومقدار محتوى الفيديو على YouTube الذي يمكن الوصول إليه.

لوحة قيادة يوتيوب

الفيديو	مصدر الزيارات	المنطقة الجغرافية	عمر المشاهد	جنس المشاهد	التاريخ	حالة الاشتراك	مصدر الاشتراكات	المزيد

المنطقة الجغرافية	عدد المشاهدات	متوسّط مدة المشاهدة	وقت المشاهدة (بالساعات)
المجموع	**39,269**	**3:36**	**2,361.3**
تونس	18.3% 7,193	4:08	21.1% 497.2
المملكة العربية السعودية	7.0% 2,749	4:57	9.6% 227.2
المغرب	4.7% 1,856	4:48	6.3% 148.5
الجزائر	3.1% 1,215	5:48	5.0% 117.7
فلسطين	1.9% 746	4:19	2.3% 53.8
مصر	1.3% 522	4:08	1.5% 36.0
قطر	0.9% 339	5:42	1.4% 32.2
فرنسا	0.5% 189	3:47	0.5% 11.9
البحرين	0.3% 129	5:43	0.5% 12.3
الإمارات العربية المتحدة	0.3% 110	3:45	0.3% 6.9

المعطيات الجغرافية للمشاهدين



رسم بياني من YouTube Analytics خلال فترة 90 يوما

من خلال الرسم البياني للمشاهدات خلال آخر 90 يومًا ما نتبين نمط مشاهدة المستخدمين. على سبيل المثال ، يظهر نمط العرض في البداية الذروة ثم ينخفض وهذا بسبب عطلات نهاية الأسبوع. يزداد تفاعل المشاهد مع الفيديو خلال أيام الأسبوع العادية ، لأنه محتوى تعليمي.

أيضا الارتفاعات الحادة والانخفاضات الحادة في المشاهدات يمكن أن تقدم معطيات مفيدة لصانع المحتوى. فربط التوجهات التي تبينها الاحصائيات بالمحتويات المرفوعة على المنصة يسمح بتحسين المحتوى.

نظرة عامة | المحتوى | الجمهور | أبحاث

المقاطع الرئيسية التي تؤثر في مقياس "الاحتفاظ بالجمهور"

أحدث الفيديوهات (آخر 365 يومًا)

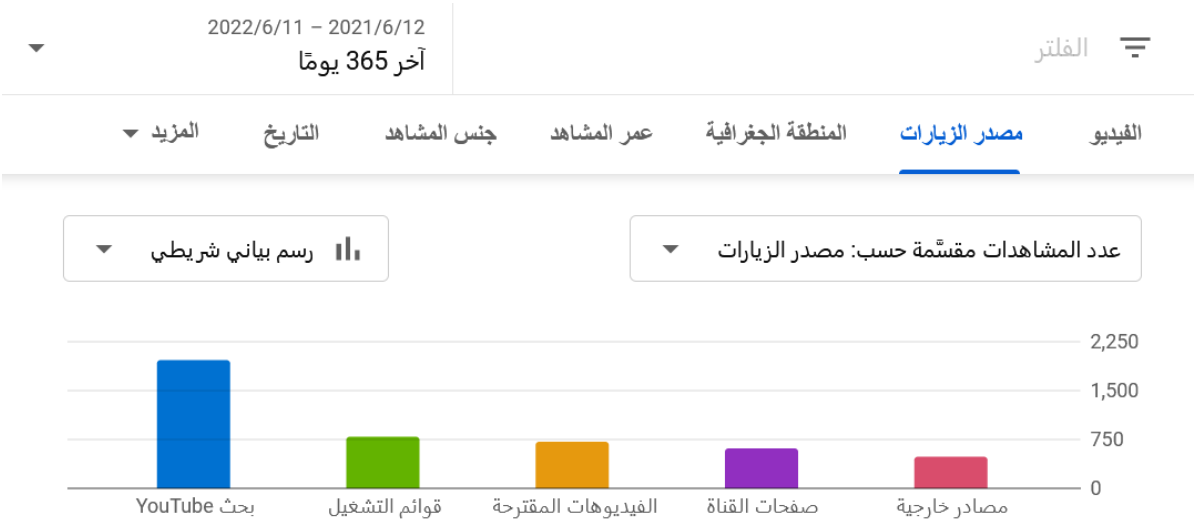
مقدّمات ذات نسبة احتفاظ بالجمهور أعلى من المعتاد	نسبة المستخدمين الذين يواصلون مشاهدة الفيديو عند الطابع الزمني 0:30
Interfaces Java	54%
Gestion des versions avec SVN sous Eclipse	54%
مقدّمات ذات نسبة احتفاظ بالجمهور أدنى من المعتاد	
Application JEE Web MVC Partie 2_03	46%
Java CheckStyle	42%

تحليلات نسبة الاحتفاظ بالجمهور

كما أشرنا سابقا من المهم اختيار الرسوم البيانية المناسبة. حيث يمكن في الكثير من الأحيان القيام بالاختيار لتسهيل عملية قراءة البيانات. كمثال, لدراسة عدد المشاهدات حسب مصدر الزيارات الرسم البياني الشريطي

أكثر وضوحا من الرسم البياني الخطي:

عدد المشاهدات حسب مصدر الزيارات من خلال رسم بياني خطي



عدد المشاهدات حسب مصدر الزيارات من خلال رسم بياني شريطي

4. 4لوحة تحليلات الموك ((**MOOCs**

في الأقسام السابقة، رأينا البيانات التي تم جمعها من موودل (Moodle) و اليوتيوب(YouTube)) والبيانات المهمة التي يجب عرضها. مثلا لو اخذنا لوحة القيادة الخاصة بتحليلات MOODLE فنجد ثلاث انواع من البيانات:

– أولا التسجيل, وتسمح بالإجابة على أسئلة على غرار: كم عدد المستخدمين المسجلين في الدورة ؟ ما هي حالة التسجيل على مدار الأسابيع؟ هل هم مستمرون؟ أم هم تركوا الدراسة؟

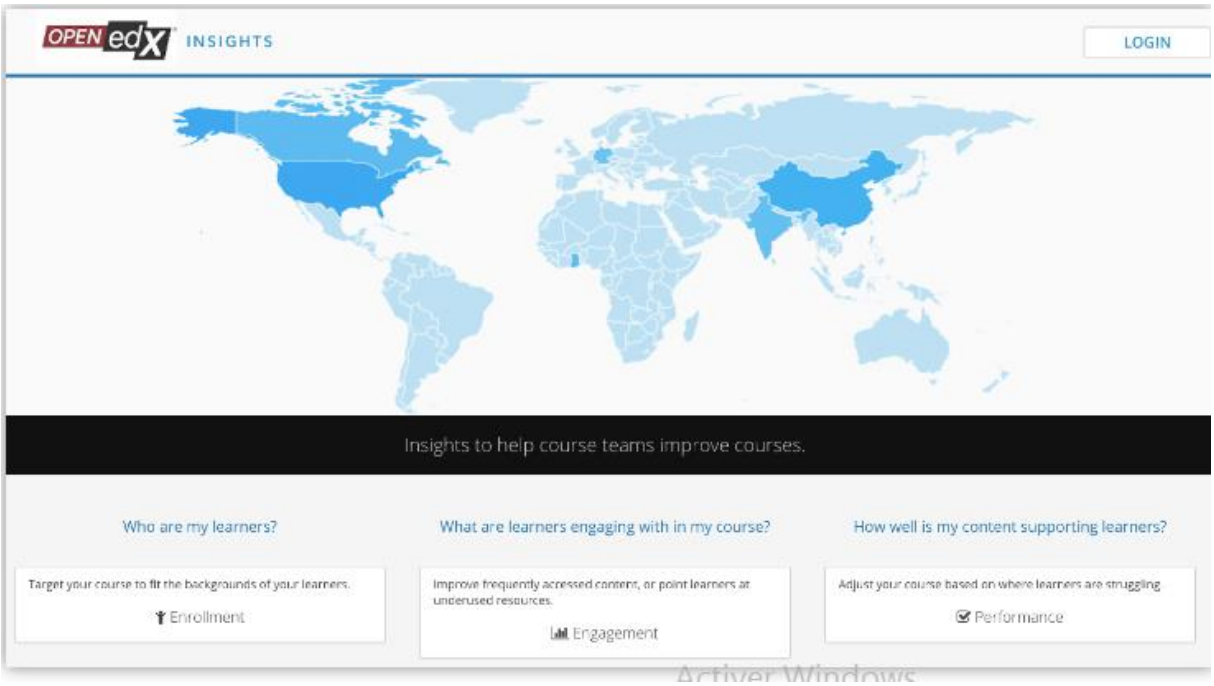
–ثانيا مشاركتهم في الدورة التدريبية ,وتسمح بالإجابة على أسئلة على غرار: هل هم منخرطون؟ كيف يتفاعلون مع المحتويات التعليمية؟

–ثالثا المعلومات الخاصة بالأداء : الاختبار البعدي و الاختبار التمهيدي.

في هذا المقطع ، سنناقش لوحة القيادة الخاصة بتحليلات الموك (MOOCs) مع الأخذ كمثال توضيحي النظام الذي يستخدمه المعهد الهندي للتكنولوجيا بومباي (IIT Bombay) على موقعه على الإنترنت.[35] يذكر المشرفون أنهم يراقبون دورات MOOCs من موقع "insights.iitbombayx.in". أولا للدخول ومشاهدة دورات الموك (MOOCs) لا بد من القيام بالتسجيل باستخدام اسم المستخدم وكلمة المرور. عند الولوج تفتح صفحة الويب التحليلات IIT Bombay edX" "والتي تتعلق بدورة تسمى "علم أصول

35 https://iitbombayx.in/courses

التدريس من أجل الاستخدام الفعال لتكنولوجيا المعلومات والاتصالات" لمعلمي المدارس. هنالك أربعة أنواع من التحليلات في لوحة قيادة باستخدام رسوم بيانية واضحة : التسجيل(Enrollment)، المشاركة (Engagement)، الأداء (Performance) والمتعلمين (Learners).

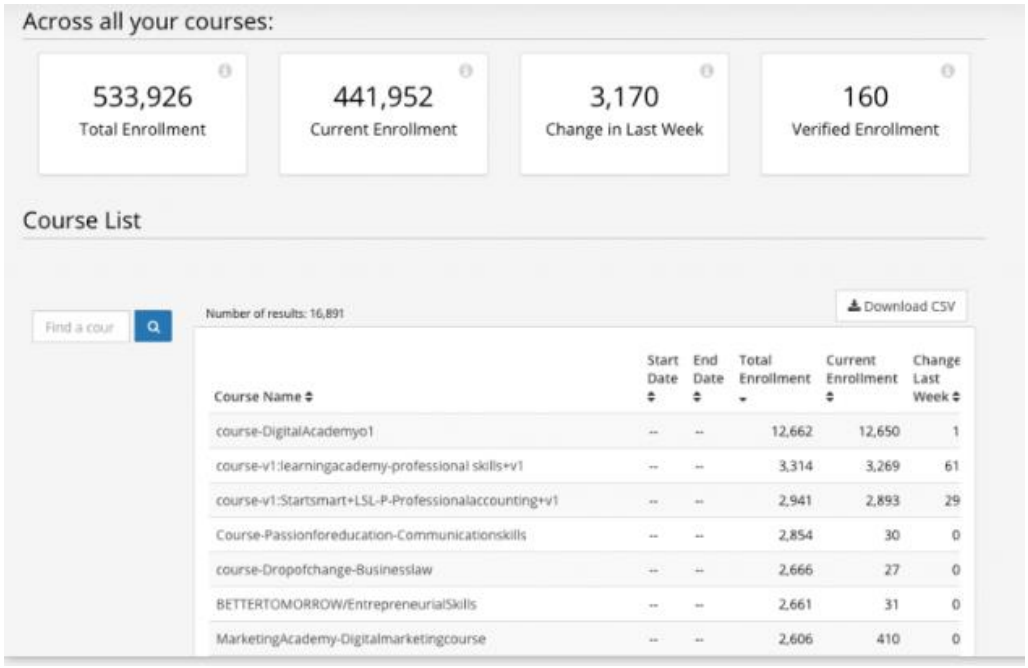


لوحة تحليلات الموك تحت منصة OPENedx للمعهد الهندي للتكنولوجيا ببومباي: الصفحة الرئيسية

إذا تحدثنا عن عمليات التسجيل فهي تقع على مدار اليوم و من الضروري أن يكون هناك تتبع كل عمليات محاولات الدخول (Tracking) . في هذه الحالة، سوف تكون لدينا معلومات ديموغرافية (كم عدد الذكور و كم عدد الإناث) وجغرافية (من أي منطقة اومن اي دولة تم الدخول إلى الموقع). نلاحظ أن التسجيل ليس بالعدد المنتظر مقارنة بعدد الطلاب الذين يتابعون الدورات لأنه يمكنهم الانتقال مباشرة إلى الدورات دون القيام بهذه العملية. تظهر هذه الحقيقة باستخدام مخطط المكدس((Stack Chart. يمكنك أن ترى عمر المستخدم ، والتعليم ، والجنس. ومع ذلك، فإن هذه البيانات اختيارية إذا قام المستخدم بإدخال هذه البيانات ، فسيتم

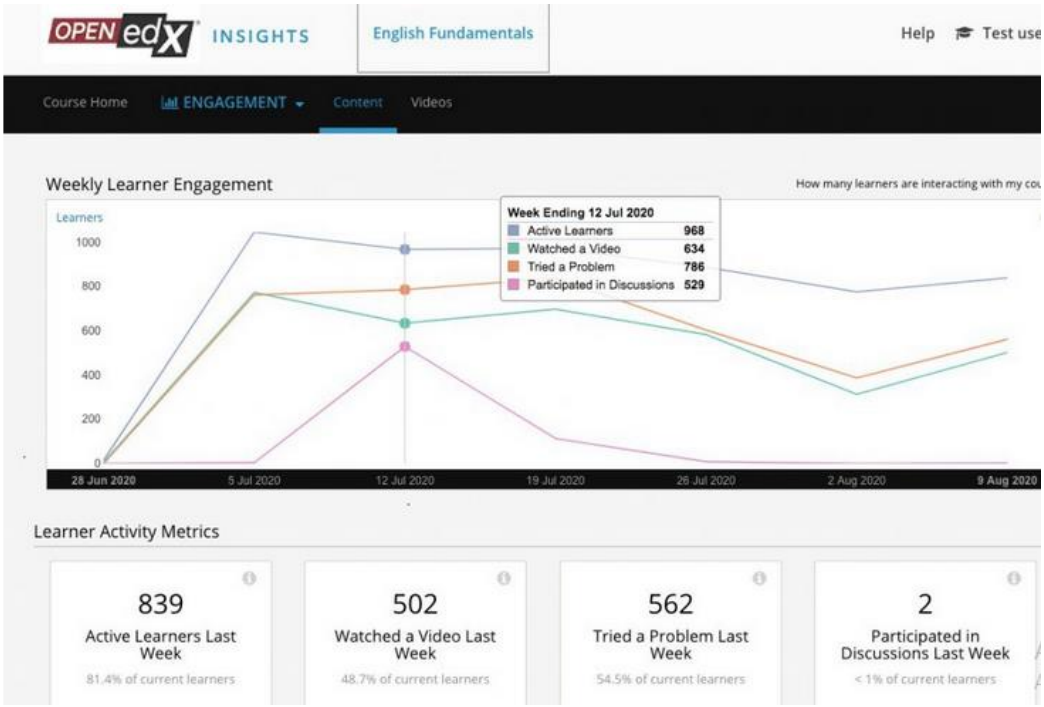
عرضها فقط وإلا فلن يتم عرضها. عندما تنظر إلى الجغرافيا ، فإن معظم المشاركين من الهند وعدد قليل من المشاركين من تنزانيا ، وهناك مشارك واحد من الولايات المتحدة.

والأهم من ذلك، يوفر OPENedX أيضًا الشكل المجدول للبيانات. لا يقتصر تصور البيانات على إظهار البيانات في رسم بياني فحسب، بل يجب أيضًا أن يحتوي على بيانات ، حتى يتمكن المستخدمون من استغلالها.



لوحة تحليلات الموك تحت منصة OPENedx للمعهد الهندي للتكنولوجيا ببومباي: واجهة المستخدم

ننتقل إلى النوع الثاني من التحليلات ، وهو يخص انخراط المتعلم. في البداية، يمكن رؤية الكثير من المستخدمين الناشطين في الدورات المقدمة والمبرمجة على اربع اسابيع، كما نلاحظ أن هذه المشاركة تنخفض تدريجياً مع تقدم

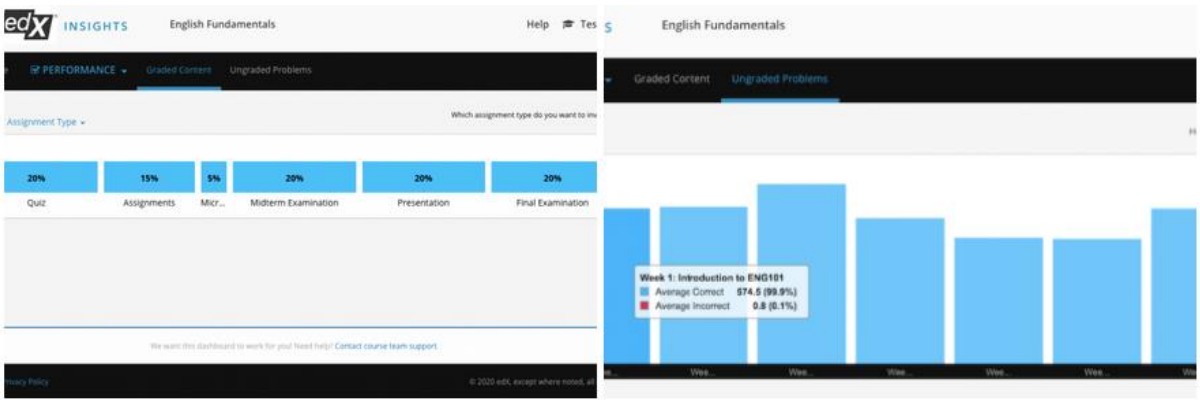


الأسابيع.

لوحة تحليلات الموك تحت منصة OPENedx للمعهد الهندي للتكنولوجيا ببومباي: واجهة انخراط المتعلم

يظهر المتعلم النشط في مخطط يوضح عدد المشكلات التي جربها وعدد المشكلات التي حاول الإجابة عليها ومدى مشاهدته لمقاطع الفيديو. نعني بالمستخدمين النشطين الذين قاموا بتسجيل الدخول إلى الفيديو، أو نظروا إلى صفحة أو مزيج من هذين الموارد أو عدد مقاطع الفيديو التي شاهدوها. من الاستعمالات المفيدة أيضا لهذه البيانات يمكن أن نذكر متابعة إحصائيات مشاهدة الفيديوهات. فمثلا ملاحظة مغادرة أغلب الطلبة لفيديو ما دون إكمال مشاهدة حتى نصف المحتوى, قد يكون مؤشر قوي لمراجعة الفيديو قصد تحسينه أو استبداله بآخر.

ننتقل إلى النوع الثالث من التحليلات الأداء: وهو يخص اختبار المعرفة واختبار التفكير ومسابقة الاستيعاب وتخصيص الموارد. هنا تسمح لوحة القيادة بعرض أداء الطلاب. من هنا يمكن تقدير مدى تطور الطالب من خلال مقارنة تطور مستواه في الإختبارات. كما يسمح بتقييم مدى تناسب الإختبارات وإن كانت تستحق المراجعة والتحسين.



لوحة تحليلات الموك تحت منصة OPENedx للمعهد الهندي للتكنولوجيا ببومباي: واجهة تقييم الأداء في دورة الإنجليزية

4. 5تحليلات التعلم التنبؤية

الهدف من التحليلات التنبؤية هو تحليل البيانات الحالية والتاريخية للتنبؤ بالأحداث المستقبلية. لذلك ، في التحليلات التنبؤية ، سنقوم بإشراك أدوات تعلم الآلة (Machine Learning) واستخراج البيانات في تحليلات التعلم الخاصة.

في تحليلات التعلم التنبؤية ، الهدف هو التنبؤ بسلوك المتعلمين وذلك قصد استباق الصعوبات التعليمية ومحاولة تخفيفها. من أجل إجراء التحليلات التنبؤية أو التنبؤ بالحدث المستقبلي لتفاعل المتعلمين، نحتاج إلى تسجيل السلوك التاريخي للمتعلمين. من خلال البيانات التاريخية يمكن إنشاء نموذج ثم استقراء هذا النموذج للتنبؤ بالأحداث المستقبلية. تتضمن التحليلات التنبؤية أيضًا تحليلات وصفية وتشخيصية ؛ على سبيل المثال ، نحتاج

أولاً إلى وصف البيانات ، ثم نحتاج إلى تحديد النمط الذي يمثل التحليلات الوصفية أو التحليلات التشخيصية ، ثم نحتاج إلى توسيع أو استقراء هذا النموذج للتنبؤ بالأحداث المستقبلية.

إذا التحليلات التنبؤية تحتاج إلى بيانات تاريخية للتنبؤ بالأحداث المستقبلية. لكن، هل نحتاج إلى بيانات من نفس الطالب أو هل يمكننا استخدام البيانات من المتعلمين الآخرين؟ لو يمكن استخدام البيانات من المتعلمين الآخرين، فما هي القيود، وما هي الافتراضات؟ من الواضح أنه لا يمكن جمع بيانات كافية عن المتعلم عند بداية استخدامه للبيئة التعليمية. في هاته الحالة، ومن أجل البدء بتقديم التوصيات والدعم، يجب الانتظار مدة كافية من الزمن. هذا غير مناسب عمليا عندما نكون في حاجة إلى تقديم العون للمتعلم فور انضمامه. لذلك في هذه الحالة يتم استخدام بيانات المتعلمين الآخرين أو استعمال بيانات من بيئة تعلم مماثلة.

تحليلات التعلم التنبئية مفيدة لمختلف المتدخلين. بالنسبة للمتعلم، تسمح تحليلات التعلم التنبؤية، بتقديم الدعم اللازم (كالاقتراحات أو التعليقات أو المساندة الإضافية) لتحسين مردوده التعليمي.

صاحب المصلحة الثاني هو المعلم. من المهم جدًا التنبؤ بما سيحدث خلال الدورة التعليمية من أجل أخذ القرارات المناسبة التي من شأنها تحسين جودة التعلم. إذ تسمح هذه التحليلات خاصة للمعلم أن يعرف متى يتحتم عليه التدخل و كيف يجب عليه أن يتدخل لتجنب الأسوء. على سبيل المثال، يمكن للمدرس معرفة أنه في الفصل من أصل 100متعلم ، قد لا يقوم 30 أو40 منهم بإرسال المهمة في الشهر المقبل. فماذا يمكن أن يفعل؟ يمكن للمدرس التفكير في استراتيجيات أخرى مختلفة وتحفيز الطلاب ومطالبتهم بإرسال المهام في الوقت المحدد أو يمكنهم تمديد الموعد النهائي للمهمة. لذا ، فإن الأنشطة المماثلة مفيدة للمعلم. بالتالي تعد تحليلات التعلم التنبؤية

مهمة جدًا للمعلمين كأصحاب مصلحة إما لمساندة المتعلمين المهددين بصعوبات التعلم قبل حتى حدوثها، وإما لتوقع المشاكل المرتبطة بالمحتوى التعليمي الحالي مما يسمح بتطويره قبل حدوث أي إشكال.

بالنسبة لأصحاب المصلحة الآخرين : المعاهد أو الأكاديميين أو رؤساء المعاهد أو رؤساء المناطق التعليمية, وغيرهم من المسؤولين, فهم معنيون أيضًا بتحليلات التعلم التنبؤية. فمثل هذه التحليلات قد تسمح بتجنب فتح مساق دراسي ما. مع ما يترتب عليه من تكلفة وجهد, إذا أكدت التحليلات التنبؤية بأن هذا المساق لن يلقى إقبالا من المتعلمين.

5. الخاتمة

في هذا الباب الثالث والأخير من هذه الدراسة فحصنا مفهوم تحليلات التعلم ، وقدمنا واقع تحليلات التعلم. تحدثنا أيضًا عن أنواع تحليلات التعلم :الوصف, التشخيص, التوقع والوصفة. كما ناقشنا بعض التفاصيل حول البيانات التي يمكن جمعها في الموك، والبيانات التي يمكن جمعها في بيئة التعليمية المعززة بالتكنولوجيا (TELE) و أيضا كيفية جمع البيانات في بيئة الفصل الدراسي.

تناولنا أيضا بالدرس قضايا الأخلاقيات والخصوصية بالنسبة للبيانات التعليمية .ثم تطرقنا إلى أدوات تحليل بيانات التعلم واستخداماتها من خلال العديد من الأمثلة. كما تناولنا بالدرس الاستخدامات الشائعة لمثل هذه التحليلات التعليمية.

الخاتمة والمقترحات

تناولت هذه الدراسة موضوع البيانات الضخمة وتحليلات التعلم. واعتنت بالتعريف بمختلف المفاهيم ذات الصلة مع ذكر تطبيقاتها. كما أنها حرصت على جعل المحتوى متاحا للمختصين وغير المختصين في مجال تكنولوجيات المعلومات والإتصال. حيث أن الهدف الأول يشمل الجانب التثقيفي مع الجانب التحسيسي. وتجدر الإشارة أنه لم يتم إغفال الجانب التقني التكنولوجي الذي تم تناوله بالتقديم لكن دون الخوض في التفاصيل التقنية. وقد تم تبويب الدراسة إلى ثلاثة أبواب, اهتم الأول منها بمفهوم البيانات الضخمة متناولا تعريفها, شارحا أسباب ظهورها, مقدما أهم الحلول التكنولوجية ذات الصلة, معددا استخداماتها. أما الباب الثاني للدراسة فقد تطرق إلى مفهوم تحليل البيانات والإمكانيات الضخمة التي يمنحها للبشرية في مختلف أوجه الحياة. كما عرج على مدى اعتماده على البيانات الضخمة معددا مواردها وأصنافها وطرق عرضها. كما تناول بالدرس أهم أصناف تحليلات البيانات. في ختام الباب الثاني تم تقديم أحد أهم أدوات التحليل الآلي للبيانات وهي خوارزميات تعلم الآلة. كما تم تقديم أنواع هذه الخوارزميات وطريقة استخدامها واستنباط المعارف منها. أما الباب الثالث والأخير فقد تطرق إلى تحليلات التعلم وهي أحد وجوه استثمار البيانات التعليمية الضخمة من خلال تحليلها واستثمارها. حيث تم تناول المفاهيم الأساسية لتحليلات التعلم مع تقديم مستوياتها الأربعة. كذلك تم الخوض في قضية جمع البيانات والتطرق إلى قضية الخصوصية وأخلاقيات جمع البيانات. كما تم تقديم أدوات تحليل البيانات واستخداماتها. فما هي أهم الأفكار المستخلصة والتوصيات ذات الصلة ؟

تحليلات التعلم هي أداة أساسية من أجل فهم وتحسين وتطوير العملية التعليمية. تنقسم هذه التحليلات إلى أصناف تتراوح من التحليلات الوصفية وصولاً إلى التحليلات التنبؤية. تسمح هذه التحليلات لأصحاب المصلحة

(من متعلمين ومعلمين وإداريين ومسؤولين ومختلف المعنيين بالعملية التعليمية) باتخاذ قرارات مستنيرة من أجل تحقيق أهدافهم الاستراتيجية.

يمكن تناول هذا الموضوع تقنيًا ، وفي هذه الحالة ، يجب على الأشخاص المعنيين بتحليلات التعلم (المدربين والفنيين وربما المتعلمين والوكلاء الإداريين) إتقان هذه المهارة من خلال اكتساب المعارف النظرية اللازمة من ناحية, وممارسات هذه المعارف من أجل استخدام مختلف الأدوات بشكل فعال لفهم سلوك المتعلم وأدائه ، ومن أجل تحسين جودة المحتوى التعليمي وغيرها من الأهداف ذات الصلة. هذه الرؤية التقنية ضرورية بما أن تحليلات التعلم هي مكون لا بد منه في أي نظام تعليمي.

الجانب التقني هو جانب تخصصي بامتياز. لكن هذه الدراسة اهتمت بالجانب التوعوي والتثقيفي أساسا مع الإشارة إلى الجوانب التقنية. فمن المهم الوعي بأهمية تحليل البيانات عموما وتحليلات التعلم خصوصا كوسيلة لتحسين أداء المتعلم ونوعية التدريس من ضمن العديد من الأهداف الممكنة. إلى جانب هذا فلم تغفل هذه الدراسة عن التطرق لاستعمال التحليلات كأداة استراتيجية من أجل حوكمة أفضل للمؤسسات التعليمية. الغرض من الحوكمة, كما هو معلوم, هو خلق قيمة مضافة لأصحاب المصلحة في المؤسسة ، من خلال تقليل المخاطر (مثل الحد من نسبة إخفاق المتعلمين) وتحسين الموارد (مثل زيادة جاذبية المؤسسة التعليمية).

يجب اعتبار تحليلات التعلم كمشروع تطوير ذو قيمة تقنية–تربوية يتوافق مع الأهداف الإستراتيجية للمؤسسة التعليمية. ومع التطور السريع لتكنولوجيات المعلومات والاتصال من جهة، واحتدام المنافسة من جهة أخرى، يجب الحفاظ على استمرارية هذا المشروع حتى تتمكن المؤسسة من تحقيق أهدافها.

لإنشاء مثل هذا المشروع، يجب على أي مؤسسة أن تخطط لعمليات التحسين المستمر وإدارة التغيير, المتعلقة بإدارة وتحليل بيانات التعلم, وذلك من خلال التكوين وتعزيز الموارد وما إلى ذلك من أدوات إدارة التغيير. في هذا الإطار من المهم أيضا أن تكون مراحل بناء المشروع قابلة للقياس و موحدة وقادرة على خلق قيمة مضافة. هنا من المهم الاعتماد على نماذج للحوكمة وعلى إطار مرجعي للممارسات الجيدة.

نلاحظ اليوم أن العديد من مؤسسات التعليم, خاصة الجامعي, قد اتخذت بالفعل خطوات مهمة نحو تحسين جودة التدريس من خلال اعتماد دبلوماتها من قبل هيئات معتمدة مثل ABET[36] و EUR-ACE[37] وغيرها. نعتقد أن اعتماد نظام مرجعي للحوكمة باستخدام تحليلات التعلم يمكن أن يمكّنها من تحقيق الأهداف المطلوبة بشكل أكثر فعالية . هنا تبرز الحاجة إلى هيكل مؤسساتي مستقل يكون هيئة اعتماد في مجال حوكمة وإدارة تحليلات التعلم في العالم العربي. هنا يمكن إما الاعتماد على المنظمات المختصة القائمة, خاصة المنظمة العربية للتربية والثقافة والعلوم بصفتها منظمة تعنى بالمجال التربوي, وذلك من خلال إنشاء لجنة دائمة تقوم بالإرشاد وتضع الشروط اللازمة لاعتماد (Accreditation) المؤسسات التعليمية في مجال تحليلات التعلم, وتقوم باعتماد المؤسسات التي تستجيب للشروط وترغب في الحصول على اعتماد رسمي. وإما يمكن إنشاء هيكل مستقل وغير ربحي متخصص في اعتماد المؤسسات التعليمية في مجال تحليلات التعلم في العالم العربي.

---

36 ABET : Accreditation Board for Engineering and Technology, Inc. - a non-governmental organization that accredits post-secondary education programs in applied and natural science, computing, engineering and engineering technology.

37 EUR-ACE :European Accredited Engineer - The EUR-ACE label (European Accredited Engineer) is managed by the ENAEE association (European Network for Accreditation of Engineering Education), which gives a mandate to national bodies to issue this certification, which must be renewed every five years.

في هذا السياق تعتبر تحليلات التعلم مشروعًا يتمحور حول مكونين الأول تقني والثاني استراتيجي. من أجل بناء هذا المشروع يجب أن تعمل على كل مكون مجموعة ذات خلفيات معرفية (نظرية وعملية) مختلفة ومتكاملة ذات علاقة بالمكون (التقني أو الاستراتيجي).

يهم المكون التقني الإداريين والمعلمين وكل من تقنيي ومهندسي الكمبيوتر الذين يعملون على نظم المعلومات التعليمية. و ليكونوا فاعلين نوصي بتأهيلهم من خلال برنامج تكويني يتناول النقاط التالية:

- توعية الأشخاص المعنيين بأهمية تحليلات التعلم في تطوير مؤسساتهم.
- تكوينهم على استخدام أدوات تحليلات التعلم الأكثر انتشارا.
- تطبيق تحليلات التعلم واختبار أدوات التحليل على مجموعة بيانات تعليمية

بالنسبة للمكون الاستراتيجي ، فيهم المشرفين والمديرين والمسؤولين وصناع القرار في مجال التعليم. في هذا السياق ، ليكونوا فاعلين نوصي بتأهيلهم من خلال برنامج تكويني يتناول النقاط التالية:

- توعية الأشخاص المعنيين بأهمية تحليلات التعلم في تطوير مؤسساتهم.
- تعريفهم بتقنيات المعلومات والاتصال الناشئة التي تدعم تحليلات التعلم: البيانات الضخمة ، Blockchain ، الأمن السيبراني ، الذكاء الاصطناعي ، التعلم الآلي ، ذكاء الأعمال ، إلخ.
- تعريفهم بمفاهيم المواءمة الاستراتيجية والحوكمة بمستوياتها الأربعة (حوكمة المؤسسة ، حوكمة نظام المعلومات ، حوكمة خدمات تكنولوجيا المعلومات ، وحوكمة البيانات).
- تدريبهم على استعمال بعض أدوات تحليلات التعلم لبناء لوحة قيادة تساعد في اتخاذ القرارات وبناء استراتيجيات في مجال التعلم.

في الختام نرجو أن تلاقي هذه الدراسة القبول والاستحسان. كما نطمح أن يساهم هذا الجهد في تطور ورقي منظومة التعليم في العالم العربي. حيث نعتقد أن المحتوى قد تناول مختلف المفاهيم والتقنيات والاستعمالات الخاصة بالبيانات الضخمة وتحليل البيانات وتحليلات التعلم. إذ أنها بمثابة النافذة المفتوحة على الآفاق الرحبة للعالم الرقمي عموما وعلى تحليلات التعلم خصوصا وما تقدمه من فرص للنهوض بالمنظومة التعليمية. التوصيات والمقترحات المقدمة بدورها تمثل جزءا هاما وأساسيا من الدراسة. نرجو أن تكون مفيدة بدورها وتمثل نقطة بداية تتلوها مبادرات أخرى من أجل مستقبل افضل للتعليم والتعلم بعالمنا العربي.

المراجع

Perret, X., Jaquemelle, G., Richard, S., "Big Data : Le Cinéma avait déjà tout prévu", Edition Kawa, 2014.

Perrichot R., "Chiffres Instagram - 2021", article publié dans le Blog Du Modérateur (BDM) en juin 2018 et actualisé en février 2021, Lien: https://www.blogdumoderateur.com/chiffres-instagram/.

IBM, Big Data-University, "BigDataAnalytics – Lesson 1:What is BigData", 2017.

Marr, B., "BigData:The 5Vs Everyone MustKnow", from the Book "Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results", Hardcover, 2016.

", King Abdelaziz University and Oum El Koura البيانات الضخمة ومجالات تطبيقهاEl Bar, A., M., Marhabe, K., A., " University, Blog: international development forum, 2018.

Teboul, B., Berthier, T., "Valeur et Véracité de la donnée :Enjeux pour l'entreprise et défis pour le Data Scientist", Actes du colloque "La donnée n'est pas donnée", École Militaire, France, 2015.

White, T., "Hadoop, The Definite Guide", O'Reilly Media editor, 2009.

Chollet, F. Deep learning with Python. Shelter Island Manning, 2nd edition, ISBN 9781617296864, 2021.

.

ISBN 978-0-387-31073-2, ، Springer،Bishop, C. M., Pattern Recognition and Machine Learning 2006.

Piryonesi, S. M., El-Diraby, T. E., "Data Analytics in Asset Management: Cost-Effective Prediction of doi:10.1061/(ASCE)IS.1943-، 26 (1)، Journal of Infrastructure Systems،the Pavement Condition Index" 555X.0000512, 2020.

Piryonesi S. Madeh, El-Diraby Tamer E., "Role of Data Analytics in Infrastructure Asset Management: Journal of Transportation Engineering, Part B: ،Overcoming Data Size and Quality Problems" doi:10.1061/JPEODX.0000175, 2020.، 146 (2): 04020022،Pavements

Farhat R. , Mourali Y., JEMNI M. and Ezzedine H. , An overview of Machine Learning Technologies and their use in E-learning, Organization of Knowledge and Advanced Technologies, OCTA'19, 2020.

Farhat R., Educational data analysis, INFOL@NGUES, 18-19-20, 2019.

Provost, F., & Fawcett, T., Data Science for Business: What you need to know about data mining and data-analytic thinking. " O'Reilly Media, Inc.", 2013.

IBM Journal of ،Samuel, A. L., "Some Studies in Machine Learning Using the Game of Checkers" ISSN 0018-8646, 1959.، doi:10.1147/rd.33.0210، **3** (3): 210–229،Research and Development

Mitchell, T. , Machine' Learning, McGraw Hill. ISBN 0-07-042807-7, 1997.

Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning, Springer. ISBN 0-387-95284-5, 2001.

Rajendran, R., "Introduction to Learning Analytics", IIT Bombay, India, 2019.

Guettat, B., Farhat, R., Jemni, M., "Activities'traces sharing for e-learners using heterogeneous LMSs", Proceedings of the International Conference on Electrical Engineering and Software Applications (ICEESA), Tunisia, 2013, appeared in IEEE-Xplore in 20/06/2016.

Tlili, A., Denden, M., Affouneh, S., Salha, S.H., Cai, Z., Jemni, M., Bozkurt, A., Huang, R., Zhu, L., "Understanding Arab Students'Behavioral Patterns in an Online Course: An Explanatory Study Based on Hofstede's National Cultural Dimensions", DOI: 10.3390/su132212426, License:CC BY 4.0, 2021.
Lang, C., Siemens, G., Wise, A., Gasevic, D., "Handbook of Learning Analytics", ed. SOLAR, 2017.

Gasevic, D., Dawson, S., Siemens, G., "Cambridge handbook of the Learning sciences". Cambridge, UK: Cambridge University Press, 2015.

Tlili, A., Altinay, F., Huang, R., Altinay, Z., Olivier, J., Mishra, S., Jemni, M., Burgo, D., " Are we there yet? A systematic literature review of Open Educational Resources in Africa: A combined content and bibliometric analysis", DOI: 10.1371/Journal.pone.0262615/, 2022.

Baker, R., & Siemens, G., "Educational data mining and learning analytics", In R. K. Sawyer (Ed.), 2014.

Siemens, G., "Learning Analytics: Envisioning a research discipline and a domain of practice", LAK'12: 2nd International Conference on Learning Analytics & Knowledge, Canada, 2012.

Siemens,G., Gašević,D., "Special Issue on Learning and Knowledge Analytics", Educational Technology & Society, 2012.

Duval, E., "Learning Analytics for Visualization and Recommendation", LAK11, Canada, 2011.

Guettat, B., Farhat, R., Jemni, M., " An Approach to Assist Learners to Build Their Own Curriculum In Personal Learning Environment Context", Indian Journal of Science and Technology, ISSN:0974-6846, 2016.

Graf, S., Ives, C., Rahman, N., Ferri, A., "A Tool for Accessing and Analyzing Students Behaviour Data in Learning Systems", LAK11, Canada, 2011.

Aditi, K., Sahana, M., "MEttLE : a Modeling-based Learning environment for undergraduate engineering estimation problem solving", Research & Practice in Technology Enhanced Learning, SpringOpen, 2022.

الدكتور رمزي فرحات هو باحث وأستاذ جامعي بالمدرسة الوطنية العليا للمهندسين بتونس بجامعة تونس. تحصل على شهادة مهندس إعلامية من جامعة المنار بتونس، ثم شهادة الدراسات المعمقة في الإعلامية من جامعة إيفري فال ديسون بفرنسا، ثم شهادة الدكتوراه في الإعلامية من جامعة إيفري فال ديسون بفرنسا وجامعة منوبة بتونس.

شمل الجانب البحثي أساسا قضايا التعليم الإلكتروني، والتعليم عن بعد باستخدام الشبكات، وبيئات التعلم الذكي، وتحليلات التعلم. في هذا الإطار قام الدكتور رمزي فرحات بتأطير العديد من طلبة ماجستير البحث وشارك في تأطير طلبة دكتوراه. وهو عضو في مخبر البحث في تكنولوجيات الاتصال والمعلومات والهندسة الكهربائية بتونس. وله ما يزيد عن 30 بحث محكم ومنشور. كما أنه ساهم كعضو تحكيم في عشرات المؤتمرات والمجلات البحثية.

الدكتور رمزي فرحات هو خبير لدى المنظمة العربية للتربية والثقافة والعلوم، وهو أيضا خبير لدى جامعة تونس الافتراضية، كما أنه عضو في اللجنة الوطنية لتأهيل برامج التكوين الهندسي في الإعلامية, كما ساهم أيضا في عدة لجان إجازة وماجستير.

الدكتور مُحَمَّد بلحسن قطاط هو باحث وأستاذ جامعي تكنولوجي بالمعهد الأعلى للدراسات التكنولوجية برادس. تحصل على شهادة مهندس إعلامية من جامعة المنار بتونس، ثم شهادة الدراسات المعمقة في الإعلامية من جامعة تونس 3، ثم شهادة الدكتوراه في الإعلامية من جامعة صفاقس.

الدكتور مُحَمَّد بلحسن قطاط متحصل أيضا على شهادة التعليم العالي المتخصصة في تكنولوجيا المعلومات الإدارية من جامعة تونس 3 والأستاذية في التصرف من جامعة تونس 3 وشهادة التبريز في الإعلامية.

الدكتور مُحَمَّد بلحسن قطاط هو باحث في مخبر البحث في تكنولوجيات الاتصال والمعلومات والهندسة الكهربائية بالمدرسة الوطنية للمهندسين بتونس منذ 2008 وقد شمل الجانب البحثي أساسا قضايا التعليم الإلكتروني، والتعليم عن بعد باستخدام الشبكات، وبيئات التعلم الذكي، وتحليلات التعلم. ولقد أصدر أكثر من عشرين منشورًا علميًا.

الدكتور مُحَمَّد بلحسن قطاط هو خبير لدى المنظمة العربية للتربية والثقافة والعلوم، وخبير لدى جامعة تونس الافتراضية، وهو أيضا خبير في البيداغوجيا الرقمية لدى وزارة التعليم العالي والبحث العلمي بالجمهورية التونسية منذ 2006. كما أنه عضو في اللجنة الوطنية لتأهيل برامج التكوين في الإعلامية

بالتعليم العالي منذ 2008, كما ساهم أيضا في عدة لجان إجازة، امتحان القبول في كليات الهندسة، ماجستير، تبريز، انتداب التكنولوجيين، انتداب محاضرين تكنولوجيين، انتداب أساتذة تكنولوجيين ولجنة الانتدابات في الإعلامية بجلس نواب الشعب التونسي (منذ 2001).

الدكتور مُحَمَّد بلحسن قطاط كان مديرًا لقسم "إدارة الأعمال" بالمعهد الأعلى للدراسات التكنولوجية برادس من 1997 إلى 2000 ، وهو مؤسس قسم تكنولوجيا المعلومات بنفس المعهد (سبتمبر 2000) وأداره الى موفى أبريل 2005.