

تحليل البيانات الضخمة ومستقبل الدراسات الاجتماعية Big data analysis and the future of social studies

جهيدة شاوش اخوان

جامعة محمد خيضر بسكرة (الجزائر)، djahida.chaouchkhouane@univ-biskra.dz

تاريخ الاستلام: 2022/05/18 تاريخ القبول: 2022/11/07 تاريخ النشر: 2022/11/14

Abstract:

Collecting field data has always been a stressful and costly work for researchers in the fields of social sciences, but what if all the data that the researcher might need is actually present and stored somewhere, and the researcher only has to identify the data he needs and start processing and analyzing it?

Through this article, we will try to identify the nature of big data and the possibility of benefiting from it in social research, and the caveats surrounding this process, and should we as specialists in these fields reconsider the traditional methodological techniques used in these sciences?

Key words: Data - big data - data exhaust - data analysis - social studies.

الملخص:

لطالما كان جمع البيانات الميدانية عملا مجهدا ومكلفا للباحثين في مجالات العلوم الاجتماعية، ولكن ماذا لو كانت كل البيانات التي قد يحتاجها الباحث موجودة فعلا ومخزنة في مكان ما وما على الباحث إلا تحديد البيانات التي يحتاجها والشروع في معالجتها وتحليلها؟

من خلال هذه المقالة سنحاول التعرف على طبيعة البيانات الضخمة وإمكانية الاستفادة منها في البحوث الاجتماعية، والمحاذير المحيطة بهذه العملية، وهل ينبغي علينا كمختصين في هذه المجالات أن نعيد النظر في التقنيات المنهجية التقليدية المستخدمة في هذه العلوم؟

الكلمات المفتاحية: البيانات - البيانات الضخمة - عوادم البيانات - تحليل البيانات - الدراسات الاجتماعية.

1. مقدمة:

تشكل البيانات الميدانية حجر الأساس الذي تقوم عليه العديد من الدراسات الاجتماعية، وتخلص بعد تحليلها وتفسيرها بشتى الطرق الممكنة إلى النتائج التي تتيح لها تقديم إضافة إلى الرصيد المعرفي والصرح العلمي القائم، ويحمل هذا الأمر من الأهمية قدرا كبيرا، حيث تتوقف قيمة البحوث الميدانية على كمية البيانات المتاحة ونوعيتها؛ فكلما كانت البيانات أكثر دقة وأكبر حجما كلما أمكن الوصول إلى نتائج أكثر قابلية للتعميم على مجمل الحالات والظروف المشابهة. ويعد جمع البيانات عملية مرهقة ومكلفة، ولهذا تشكل البيانات هاجسا كبيرا للباحثين عموما وخاصة منهم طلبة الدراسات الأكاديمية، الذين يستعصي عليهم الوصول إلى عينات كبيرة أو بيانات على مستوى كبير من الدقة والمصادقية.

ولقد كان من نتائج التطور الكبير في العقود الأخيرة للتقنيات الحديثة لوسائل الاتصال وانتشار الشبكة العنكبوتية إلى أوسع مجالاتها، وما صاحبها من تطور الذكاء الصناعي وعلوم الآلة وتقنيات البيانات...، كل هذا أفضى إلى تحقق أحد أهداف العلماء منذ زمن طويل وهو الحصول على كميات هائلة من البيانات المتنوعة في كل المجالات وبشكل سريع جدا عجزت أمامه أقوى أنظمة التخزين المتاحة عن استيعاب هذا الكم الهائل من البيانات. وهو ما سمي بالبيانات الضخمة.

لقد باتت البيانات الضخمة تنبئ بثورة في عالم البيانات، إذ توصف بأنها منجم للبيانات، وأنها نطف المستقبل، وقد أصبحت محاولات الاستفادة منها عاملا مشتركا بين مختلف المؤسسات في شتى المجالات. ومن هنا جاءت هذه الدراسة للبحث عن إجابة للتساؤلات التالية:

ما طبيعة هذه البيانات الضخمة؟ وكيف يمكن الاستفادة منها في البحوث الاجتماعية؟ وما حدود الاستفادة الممكنة؟ وما هي الأدوات والآليات التي تمكننا من ذلك؟ وهل يمكن أن تحدث تغييرا جذريا في تقنيات جمع وتحليل البيانات التقليدية؟...

2. ماهية البيانات الضخمة:

1.2 مفهوم البيانات Data:

البيانات أو المعطيات هي سلسلة غير مترابطة من الحقائق الموضوعية، التي يمكن الحصول عليها عن طريق الملاحظة، أو عن طريق البحث والتسجيل. وبشكل عام فالبيانات

هي مجموعة من الحروف أو الكلمات أو الأرقام أو الرموز أو الصور (الخام) المتعلقة بموضوع معين. مثال على ذلك: بيانات الموظفين (الأسماء - الأرقام الوظيفية - المهن - الصور) بدون ترتيب، وينتج عن هذه البيانات بعد المعالجة ما يطلق عليه مصطلح معلومات (موسوعة عريق).

وهي قابلة للمعالجة بطرق مختلفة لتتحول بعد ذلك إلى معلومات قابلة للاستخدام. فالبيانات إذن هي الصورة الخام للمعلومات قبل عمليات الفرز والترتيب والمعالجة ولا يمكن الاستفادة منها بصورتها الأولية قبل المعالجة.

ويمكن القول أن البيانات هي الأساس أو الوقود لمختلف العلوم التي أصبحت تحظى بنجاح غير مسبوق بفضل كمية البيانات ومصادرها الموجودة اليوم نتيجة للتحول الرقمي والاعتماد على الخدمات الإلكترونية والأنظمة الذكية واستخدام مواقع التواصل الاجتماعي بشكل أساسي ويومي، والتوجه نحو الشراء الإلكتروني، ولهذا فإن حجم البيانات قد تضخم بشكل غير مسبوق، حيث تشير الإحصائيات إلى أن 90% من البيانات المتوفرة اليوم تم إنتاجها خلال السنوات القليلة الماضية، وهي في زيادة مستمرة (عبدالسلام، 2021، صفحة 13).

2.2 البيانات الضخمة Big Data:

تختلف وتتنوع التعريفات والمفاهيم لهذا المجال ما بين الخبراء والشركات والمنظمات المتخصصة، حيث يعرف معهد ماكنزي العالمي البيانات الضخمة، أنها مجموعة من البيانات التي يفوق حجمها القدرة على معالجتها باستخدام أدوات قواعد البيانات التقليدية، من النقاط ومشاركة ونقل وتخزين وإدارة وتطيل، في غضون فترة زمنية مقبولة (مركز الإحصاء، د.ت، صفحة 4)

في تقرير بحثي وعدد من المحاضرات المتعلقة به عام 2001، قام "دوغ لاني" محلل مجموعة META Group المعروفة الآن باسم (Gartner بتعريف تحديات نمو البيانات وفرصها كعنصر ثلاثي الأبعاد، بمعنى زيادة الحجم (كمية البيانات)، السرعة (سرعة البيانات الصادرة والواردة) والتنوع (تنوع أنواع البيانات ومصادرها). وتقوم Gartner

والكثير من الشركات في هذه الصناعة الآن بالاستمرار في استخدام نموذج "3 Vs" لوصف البيانات الضخمة (منصة المستقبل العلمي).

وفي 2012، قامت Gartner بتحديث تعريفها ليصبح كالتالي: "البيانات الضخمة هي الأصول المعلوماتية كبيرة الحجم وسريعة التدفق وكثيرة التنوع، التي تتطلب طرق معالجة مجدية اقتصاديا ومبتكرة من أجل تطوير البصائر والمساعدة على اتخاذ القرارات" (Laney, 2013)

وحسب شركة (IBM) تنشأ البيانات الضخمة عن طريق كل شيء من حولنا في كل الأوقات، كل عملية رقمية وكل تبادل في وسائل التواصل الاجتماعي ينتج البيانات الضخمة، التي تتناقلها الأنظمة وأجهزة الاستشعار والأجهزة النقالة. البيانات الضخمة لها مصادر متعددة تختلف من حيث السرعة والحجم والتنوع. لكي نستفيد من البيانات الضخمة، نحتاج إلى معالجة مثالية وقدرات تحليلية ومهارات (Miele, 2013)

حسب الإتحاد الدولي للاتصالات (ITU) يشير مصطلح البيانات الضخمة إلى مجموعات البيانات التي تتميز أنها فائقة حجما وسرعة أو تنوعا، بالقياس إلى أنواع مجموعات البيانات معهودة الاستخدام (ITU, 2013)

تعريف TBDI للبيانات الضخمة: البيانات الضخمة هو مصطلح ينطبق علي الأجسام الضخمة للبيانات التي تتنوع في طبيعتها سواء أكانت منظمة، غير منظمة أو شبه منظمة، بما في ذلك من المصادر الداخلية أو الخارجية للمنظمة، ويتم توليدها بدرجة عالية من السرعة مع نموذج مضطرب، والتي لا تتفق تماماً مع مخازن البيانات التقليدية والمنظمة وتتطلب نظام إيكولوجي قوي ومعقد مع منصة حوسبة عالية الأداء وقدرات تحليلية للالتقاط ومعالجة وتحويل وكشف واستخلاص القيمة والرؤى العميقة في غضون وقت زمني مقبول (منصة المستقبل العلمي).

من خلال كل ما سبق يمكن التأكيد على ثلاث خصائص أساسية للبيانات الضخمة وهي ما اصطلح على تسميته 3(Vs):

1. الحجم ان يكون حجم البيانات ضخمة ويقاس حجمها بالبيتا بايت أو الاكسابايت.
2. السرعة وهو سرعة تدفق البيانات وتداولها.
3. التنوع اشكال مختلفة من البيانات في شكل نصوص.

و تقسم البيانات الضخمة عموما الى ثلاثة انواع:

- **بيانات منظمة أو مهيكلة:** وهي تمثل جزءا صغيرا (أقل من 10%) من البيانات منظمة في جداول أو قواعد بيانات. وتتميز بإمكانية البحث فيها وتحليلها، كما يمكن إدارتها باستخدام لغات برمجة خاصة.
- **بيانات غير منظمة:** تشكل النسبة الأكبر من البيانات، وهي البيانات التي يولدها الأشخاص يوميا من كتابات نصية وصور وفيديو ورسائل ونقرات على مواقع الانترنت... الخ. ورغم أن هذه الأنواع من البيانات لها هيكل داخلي خاص، لكنها تعتبر غير منظمة لأن بياناتها لا تتسق في موحدة مناسبة لقاعدة بيانات.
- **بيانات شبه منظمة:** وهي تقع بين النوعين السابقين، إذ تعتبر نوعا من البيانات المنظمة إلا أن البيانات لا تكون في صورة جداول أو قواعد بيانات.

3. مصادر الضخمة الكبيرة:

في الواقع نحن من يقوم بإنتاج البيانات الضخمة حتى دون أن نعرف ذلك: نصوص، صور، على الشبكات الاجتماعية، اليوتيوب، تحركاتنا على GPRS ، التعاملات البنكية التجارية... (لطابي، 2019، صفحة 60) إن البيانات الضخمة هي بيانات رقمية يتم إنشاؤها باستمرار من قبل سكان العالم. وجزء كبير من هذا الناتج هو "استفاد البيانات" أو السجلات، تم إنشاؤه كمنتج ثانوي للتفاعلات اليومية مع المنتجات أو الخدمات الرقمية، إذ يتم جمعها بشكل متزايد عن طريق أجهزة استشعار المعلومات المتنقلة، والتقنيات الحسية الجوية (الاستشعار عن بعد)، وسجلات البرامج، والكاميرات، والميكروفونات، وأجهزة تحديد نذبذبات الإرسال (تحديد الهوية باستخدام موجات الراديو وشبكات استشعار اللاسلكية) (www.marefa.org)

منذ أكثر من خمسة عقود اكتسحت تكنولوجيا المعلومات جميع مجالات الحياة (المنازل، المستشفيات، الوزارات،...). لقد أدى هذا الاستخدام القوي لأجهزة الكمبيوتر إلى إنتاج كميات كبيرة من البيانات، ومثال ذلك الشركات ذات الحجم البشري مثل Google وMicrosoft، هذه الشركات الكبيرة التي أصبح لديها مليارات البيانات للاحتفاظ بها. مثال

آخر هو شركات الهاتف التي لديها كميات كبيرة من البيانات عن العملاء والخدمات المقدمة لهم. (BAMBA, p. 3)

ويدير القطاع الخاص وخصوصا رعاية الهاتف المحمول وشركات بطاقات الائتمان ومواقع شبكات التواصل الاجتماعي مجموعات بيانات هائلة تحمل رؤى ثرية. ففي دقيقة واحدة على الإنترنت:

- 470.000 تغريدة على تويتر
 - 2 مليون لقطة مشتركة على سنابشات
 - 4 ملايين عملية بحث تم إجراؤها على Google
 - و في عالم التجارة: تنتج Walmart 40 بيتابايت يوميا من شبكات اجتماعية متعددة المصادر وتطبيقات الهاتف المحمول وأجهزة استشعار مختلفة (مراقبة ومراقبة النشاط). أنظمة الموقع ، وعلامات RFID ، إلخ. (Allemand, 2018, p. 3)
- علما وأنه يمكن متابعة إحصاءات الشبكة العالمية للمعلومات أنيا من خلال الموقع:

www.internetlivestats.com

ويزداد حجم البيانات في العالم زيادة مطردة؛ ففي عام 2020 تم إنشاء 64.2 زيتابايت من البيانات، أي بزيادة قدرها 314 بالمائة عن عام 2015، كما ساهمت زيادة الطلب على المعلومات بسبب وباء كوفيد-19 في نمو أعلى من المتوقع. والجزء الأكبر من تلك البيانات هي ما يمكن أن يُسمى "عوادم البيانات"، وهي البيانات التي تُجمع من حصيلة التفاعلات اليومية مع المنتجات والخدمات الرقمية، بما في ذلك الهواتف الخلوية وبطاقات الائتمان ومنصات التواصل الاجتماعي. فالبيانات تنمو بسبب تزايد تجميعها الهواتف الخلوية المتنوعة والرخيصة بما لها من قدرة على التقاط المعلومات. ومنذ ثمانينيات القرن الماضي، تتضاعف السعة العالمية لخرن البيانات في كل 40 شهرا. (un.org)

ومع تطور تقنيات التعرف على الوجه والأشخاص فإنها ستتمكن من العثور على المزيد من التفاصيل والمعلومات عن أي شخص، ومع تزايد عدد الأجهزة المتصلة بالإنترنت، الأجهزة التي لم نعتد عليها أن تتصل بالشبكة العالمية مثل السيارات والبرادات والغسالات... فإنها كلها تسهم في زيادة حجم البيانات المنتجة (لطاي، 2019، صفحة 62). ومن الأمثلة تقع تحت تعريف البيانات الضخمة:

- بيانات تداول أسهم البورصة.
 - بيانات وكالة الفضاء ناسا.
 - بيانات مواقع التواصل الاجتماعي.
 - بيانات سلوك مستخدمي مواقع المتاجر الإلكترونية.
 - بيانات مستخدمي الأجهزة الذكية مثل الساعات الذكية.
- الأمثلة السابقة تولد يومياً بيانات حجمها يتراوح بين 1 تيرابايت إلى 500 تيرابايت.
- عادم البيانات: البيانات المجمعّة بشكل سلبي المستمدة من الاستخدام اليومي للأجهزة الرقمية (United Nations Global Pulse, 2013, p. 1)

تقوم مؤسسات متخصصة في البحث عن المعلومات الخاصة واقتنائها ثم بيعها، بحيث يشكلون وسطاء أو سماسرة بيانات (data broker)، إن هؤلاء يعرفون الكثير عن حياتنا الخاصة، هواياتنا، الأفلام المفضلة لدينا، الرياضات التي نمارسها، الأمراض التي أصبنا بها والأعراض التي نعاني منها، مناسباتنا الاجتماعية، تنقلاتنا، علاقاتنا الاجتماعية، أرقام البطاقات الخاصة بنا...، إنهم يشتررون هذه البيانات من الشركات الكبيرة مثل جوجل، فايسبوك وتويتر... وغيرها

4. البيانات الضخمة الامتيازات والآفاق المفتوحة:

إن الهدف الأساسي من كل هذه المعلومات هو ما يمكننا استخراجها منها من معلومات والاستفادة منها في اتخاذ قرارات من خلال المعالجة والتحليل الضخم للبيانات، وذلك من خلال خوارزميات رياضية يتم إيجادها خصيصاً لتحليل هذه البيانات، وهو ما يساعد مثلاً في عمليات التسويق المستهدفة بدقة على الأنترنت؛ فالأشخاص الذين يبحثون مثلاً عن منتج معين عبر الأنترنت يكونون مستهدفين فيما بعد بإعلانات دقيقة عن نفس المنتج الذي يبحثون عنه، إلى غير ذلك من إمكانيات تطبيقه في جميع المجالات التعليمية الطبية البيئية التجارية العسكرية... (لطابي، 2019، صفحة 61) لاشك أن إمكانيات الاستفادة من بيانات موجودة بالفعل في مؤسسة ما على نحو أفضل يمثل ميزة كبيرة لأي

مؤسسة ، وربما كانت فرص ابتداع أفكار ورؤى جديدة من خلال تحليلات البيانات الضخمة أكثر وضوحا في القطاعات التي تتمتع بكثافة البيانات، مثل الرعاية الصحية والمرافق العامة والنقل والتعليم والمالية والعلوم والحكومة، وبتعزيز فهمها لعملائها ومستخدمي خدماتها يمكن لهذه المؤسسات أن تقدم خدمات أفضل وأكثر كفاءة (ابراهيمى، 2019، صفحة 49).

يتمثل دور تحليل البيانات الضخمة في التنبؤ أو ما يعرف باستشراف المستقبل، وهو التنبؤ بما سيحدث مستقبلا، مما يساعد المؤسسات في تفادي المشكلات والاستعداد لها من خلال اتخاذ القرارات المناسبة والتي تتلاءم مع الأحداث المستقبلية القادمة (شيكور، 2019، صفحة 77). حيث تمثل البيانات شريان الحياة لعملية اتخاذ القرارات، ولهذا أصبح تحليل البيانات الضخمة أمرا شائعا في القطاع الخاص اليوم، مثله في ذلك مثل عملية تمييز المستهلك وقولبة الخدمات لتتناسب الحاجات الشخصية للأفراد، فضلا عن استخدام التحليل التنبؤي في التسويق والإعلان والإدارة... وتقوم الشركات بتحليل هذه البيانات لدعم اتخاذ القرار أو تقديم معلومات عن السوق، ومؤخراً بدأت مؤسسات القطاع العام في الاستفادة من تقنيات مماثلة لتوليد رؤى قابلة للتنفيذ لصانعي السياسات .

وتقدم البيانات الضخمة بشكل عام العديد من المزايا والإمكانيات التي توفر الوقت والجهد والتكلفة العالية المطلوبة في معالجة وتحليل البيانات والتي تتلخص أهمها فيما يلي:

- القدرة على تحليل بيانات وسائل التواصل الاجتماعي لاستخراج الآراء المختلفة مما يتيح إمكانية تحليل انطباعات الجمهور تجاه قضايا ومسائل معينة لقياس مدى رضاهم من عدمه.
- التحليلات التنبؤية.
- تحليل وتقييم المخاطر المالية
- تصنيف المستخدمين وتحديد الفئة المستهدفة لموضوع معين مما يساعد على تحسين أداء وفاعلية الحملات التسويقية والإعلانية مثلا.
- تحسين خدمة وإدارة العمليات.
- كشف الاحتيال...

ويمكن اعتماد تقنيات مشابهة للحصول على بيانات آنية فيما يتعلق بمؤشرات جودة حياة الأفراد ورفاهية المجتمعات ومختلف ظروف حياتهم واتجاهاتهم واحتياجاتهم، ومن ثم إمكانية استخدامها في تقديم المعونات للفئات الضعيفة والمحاجة لثنى أنواع الدعم. ويمكن للمصادر الجديدة للبيانات والتكنولوجيات والمناهج التحليلية الجديدة إذا طبقت جميعها بشكل إيجابي أن تجعل من عملية اتخاذ القرارات مسألة أكثر كفاءة ومرونة.

ففي عام 2015 مثلا شرع العالم في العمل على جدول أعمال التنمية، الذي عُرف بأهداف التنمية المستدامة. ويتطلب تحقيق هذه الأهداف عملا متكاملًا على التحديات الاجتماعية والاقتصادية والبيئية مع التركيز على التنمية التشاركية والشاملة التي لا يتخلف عنها أحد. وقد ظهر هناك نقص في البيانات الحاسمة التي لا بد منها لوضع سياسات إنمائية وطنية وإقليمية وعالمية. إذ تفتقد كثير من الحكومات إلى سبل الحصول على البيانات الكافية المتصلة بشعوبها. وينطبق ذلك على الأفراد الأشد فقرا والأكثر تهميشا، وهم الفئات التي ينبغي على المسؤولين السعي إلى التركيز عليهم لتحقيق القضاء على الفقر المدقع والقضاء على الانبعاثات الخطيرة على البيئة مع حلول عام 2030 من خلال عملية تشمل الجميع. (un.org)

ويمكن للبيانات الضخمة أن تسلط الضوء على أوجه التفاوت في المجتمع مما لم يكن مشاهدا في السابق. فعلى سبيل المثال، تعاني النساء والفتيات (وهن غالبا ما يعملن في قطاعات غير رسمية أو في المنازل) من القيود الاجتماعية المفروضة على حركتهن، فضلا عن تهميشهن في عملية اتخاذ القرارات الخاصة منها والعامّة.

ويجمع القطاع الخاص معظم البيانات الضخمة التي تنطوي على أهمية كبيرة لاستخدامها في للصالح العام. ولذا، يُرجح انتشار الشراكات بين القطاعين العام والخاص. وسيكون التحدي المائل هو ضمان استدامتها على مر الزمن، فضلا عن ضرورة وجود أطر واضحة تحدد الأدوار التي يضطلع بها كل طرف والتوقعات المرجوة منه (un.org).

6. توسع مجالات استخدام البيانات الضخمة:

البيانات هي الشريان الحيوي لكثير من العلوم ولهذا فإن إمكانية استيعاب البيانات بأحجام أكبر مما سبق ومعالجتها في وقت أسرع سيزيد من الاستفادة وتحقيق الأهداف المرجوة من تلك العلوم، خصوصاً في ظل ازدهار مجالات الذكاء الاصطناعي. لقد أصبح من المتاح الوصول إلى كمية ضخمة جداً من البيانات المنتظمة وغير المنتظمة التي يمكن تحليلها حسابياً للحصول على نتائج تقييمية دقيقة إلى حد كبير للعديد من المؤشرات والمتغيرات، إلى جانب سهولة استخراج العلاقات والروابط بين مختلف المتغيرات خصوصاً تلك التي تتعلق بالسلوك والتفاعل البشري، مما يسمح باكتشاف ارتباطات توضح بدقة كبيرة الاتجاهات التجارية والسياسية والاجتماعية...، ويتيح هذا الأمر كذلك المساهمة الفعالة في العديد من المجالات من خلال إمكانية التنبؤ، أو الحصول على الاستشهادات والأدلة القانونية، مكافحة الجريمة... وغيرها (www.marefa.org).

ومن خلال ماسبق يتضح جلياً أنه يمكن للمشاهدات الحاصلة من مناجم المعلومات تلك أن تكمل الإحصاءات الرسمية وبيانات الاستبيانات التي يجمعها الباحثون والمختصون، مما يضيف رصيذاً هائلاً للمعلومات المتصلة بالسلوك الإنساني والاجتماعي عموماً. ويتكامل هذه البيانات الجديدة مع البيانات التقليدية، يصبح من الممكن إنتاج معلومات أنية أعلى جودة وأكثر تفصيلاً وأوثق صلة.

وهو ما أمكن استغلاله بنجاح منقطع النظير في العديد من النماذج الواقعية في مختلف المجالات، وذلك نظراً للميزات التنافسية الكبيرة التي يمكن أن تحققها، حيث تستخدم في مختلف المجالات خاصة الاقتصادية والاجتماعية من أجل تحسين آليات عمل المنظمات وتطوير أدائها والوصول إلى الفئات المستهدفة المناسبة للمنتج أو الخدمة المقدمة وتحسين نتائج عمليات التسويق من خلال رصد رغبات المستهلكين وتفضيلاتهم وسلوكهم الاستهلاكي. كما تستخدم أيضاً في مجالات البحث العلمي المختلفة (الاقتصاد، الطب، الطاقة، علم الفلك، ...)

في عام 2012، أعلنت إدارة أوياما عن مبادرة التنمية وبحوث البيانات الضخمة والتي تناولت كيفية استخدام البيانات الضخمة لمعالجة المشاكل الهامة التي تواجه

الحكومة وقد تألفت المبادرة من 84 برنامج مختلف للبيانات الضخمة موزعة علي 6 إدارات ووكالات فيدرالية تودع أكثر من 200 مليون دولار لمشاريع البيانات الضخمة البحثية.

وقد تضمنت المبادرة " National Science Foundation بعثات في الحوسبة" والتي منحت 10 ملايين دولار على مدى 5 سنوات لمعمل AMPLab ، كما تلقي AMPLab أيضاً تمويل من DARPA ، وأكثر من اثني عشر راعياً صناعياً، ويستخدم البيانات الضخمة لمواجهة مجموعة واسعة من المشاكل بدءاً من الاختناقات المرورية وحتى مكافحة السرطان.

وقد شملت مبادرة البيت الأبيض أيضاً التزاماً من وزارة الطاقة لتوفير 25 مليون دولار علي مدار 5 سنوات لإنشاء معهد إدارة وتحليل وتصوير البيانات (SDAV) ، والذي يتم قيادته من قبل معمل لورانس بيركلي الوطني التابع لوزارة الطاقة. ويهدف معهد SDAV جمع الخبرات من 6 مختبرات وطنية و 7 جامعات لتطوير أدوات جديدة لمساعدة العلماء في إدارة وتصوير البيانات علي أجهزة الكمبيوتر العملاقة الخاصة بالإدارة.

هذا وقد أعلنت ولاية ماساشوستس الأمريكية عن مبادرة ماساشوستس للبيانات الضخمة في مايو 2012، والتي توفر التمويل من حكومة الولاية وشركات القطاع الخاص لمجموعة متنوعة من المؤسسات البحثية. وقد استضاف معهد ماساشوستس للتكنولوجيا مركز إنتل للعلوم والتكنولوجيا الخاص بالبيانات الضخمة في مختبر MIT لعلوم الكمبيوتر والذكاء الاصطناعي (wikipedia).

والجدير بالذكر، أن تحليل البيانات الضخمة قد لعب دوراً كبيراً في حملة إعادة انتخاب باراك أوباما الناجحة عام 2012. وتمتلك الحكومة الاتحادية للولايات المتحدة 6 من أصل 10 أجهزة كمبيوتر تُعد هي الأكثر نفوذاً في العالم والتي يُطلق عليها " Supercomputers". وتقوم وكالة الأمن الوطني الأمريكي حالياً ببناء مركز بيانات يوتاه " Utah Data Center"، والذي سيكون قادر علي التعامل مع معلومات تقدر مساحاتها بـ يوتابايت والتي جمعها وكالة الأمن القومي عبر الإنترنت (wikipedia) .

وتقوم المفوضية الأوروبية علي مدار عامين بتمويل منتدى القطاعين العام والخاص للبيانات الضخمة من خلال برنامجهم السابع لإشراك الشركات والأكاديميات وغيرهم من

أصحاب المصلحة في مناقشة قضايا البيانات الضخمة. ويهدف المشروع إلي تحديد إستراتيجية خاصة بالبحث والابتكار لتوجيه إجراءات الدعم من المفوضية الأوروبية للتنفيذ الناجح لاقتصاد البيانات الضخمة. وسوف تستخدم نتائج هذا المشروع كمدخل لمشروعهم التالي. Horizon 2020. (wikipedia).

وهناك العديد من الأمثلة والتطبيقات التي نجحت في الاستفادة من تقنية البيانات الضخمة في عدة مجالات من أهمها (عبدالسلام، 2021، صفحة 24):

- **التعليم:** يعتبر تطبيق دالينجو لتعلم اللغات واحد من التطبيقات والأنظمة التي استطاعت الاستفادة من البيانات الكبيرة عن طريق تخزين البيانات للمستخدمين وتفاعلاتهم وتحليلها باستمرار لمعرفة نجاحات المستخدمين وإخفاقاتهم وأنشطتهم، والعمل على التحسين المستمر لذلك، يعتبر اليوم واحداً من التطبيقات المعروفة في تعليم اللغات
- **المبيعات:** استطاعت شركة أمازون معالجة ملايين العمليات كل يوم والاستفادة من تحليل السلوك الاستهلاكي لعملائها وتحديد احتياجاتهم واقتراح المنتجات والبضائع عليهم مما ساهم في زيادة المبيعات والأرباح .
- **الترفيه:** استطاعت منصة نتفليكس من استغلال وتحليل بيانات المستخدمين في تحديد المواد المناسبة واقتراحها عليهم حسب الوقت المناسب لهم
- **النقل والمواصلات:** استطاعت شركة أوبر من تحليل بيانات الركاب في تقديم خدمات وعروض تناسب احتياجاتهم وقدراتهم المالية
- **الصحة:** نجحت شركة مايكروسوفت بالتعاون مع أحد المعاهد الهندية في حصولها على كم بيانات ضخمة من سجلات، لأكثر من مليون شخص، في مساعدة الأطباء؛ في التعرف بشكل دقيق على أسباب انتشار أمراض العيون، إضافة إلى مساعدة الأطباء على تحديد الوقت المطلوب للمريض حتى تستقر حالته الصحية ويعالج بصره، وكذلك القيام بالعمليات الجراحية للعيون بنجاح
- **الخدمات المالية:** نجحت شركة أمريكان اكسبريس في استغلال قاعدة بيانات تحتوي على كمية هائلة من البيانات لمالكي البطاقة الإئتمانية من مختلف أنحاء العالم وفهم سلوك المستهلك بشكل أعمق والتنبؤ بالمعاملات المستقبلية .

- الصناعة: تعتمد شركة جنرال إلكتريك على ابتكار منتجات وطرح حلول رقمية ترتبط باحتياجات ومتطلبات المستهلكين .
- الدعاية والتسويق: يعتبر الفيسبوك واحداً من أكبر مصادر البيانات الضخمة وقد نجح باستغلالها والاستفادة منها في تصميم خوارزميات تعرض للمستخدمين إعلانات لبضائع وخدمات تناسب احتياجاتهم واهتماماتهم .

7. البيانات الضخمة كتقنية مستقبلية في الدراسات والبحوث الاجتماعية:

يرى سيث ستيفنز - دافيدوتس أنه في هذا العصر التقني الذي تقترب فيه المسافات، يقبع الناس هائمين أمام شاشاتهم يبوحن بخصوصياتهم دون أدنى قيد أو حرج، حيث أن البيانات الضخمة التي توفرها التكنولوجيا الحديثة تزيح الستار عن جوانب خفية، وحساسة، وسرية لمستخدمي الإنترنت. يتحدث دافيدوتس عن البيانات الضخمة كمستقبل الوسيلة أو كوسيلة المستقبل؛ فكرتان يراوح بينهما خلال سرده العديد من القصص والدراسات المتعلقة بالبيانات الضخمة؛ فهل ستكون الوسيلة التي من خلالها سيتمكن البشر من دراسة الماضي وفهم الحاضر واستقراء المستقبل؟ وقد احتوى كتابه "الكل يكذب" على موضوعات متعددة قام الباحث بدراستها اعتماداً على البيانات الضخمة كونه متخصصاً في علم النفس ومبرمجاً في الإعلام الآلي ما أمكنه من القيام بالعديد من الدراسات الميدانية التي اعتمد فيها كلية على تحليل مجموعات من البيانات الضخمة التي شملت موضوعات حساسة يصعب أن يجيب عليها المبحوثون بشكل صريح بالأساليب التقليدية لجمع البيانات كالاستمارة مثلاً، هذه الأخيرة التي يرى دافيدوتس أن غالبية البيانات التي نتحصل عليها من خلالها هي بيانات غير دقيقة حيث يميل المبحوثون إلى الكذب أو إخفاء الحقيقة أو ادعاء المثالية في إجاباتهم سواء كان ذلك بشكل مقصود أو غير مقصود، وهذا مصدر تسمية كتابه (الكل يكذب)، ولكن الأشخاص حين يخلون إلى هواتفهم فإنهم يبوحن بكل أسرارهم ومخاوفهم وآمالهم دون تحفظ، وهو ما يجعل البيانات التي يتم تخزينها نتاجاً للمخلفات الإلكترونية التي تتجم عن تفاعل البشر مع أجهزتهم الذكية أكثر مصداقية في نظره من البيانات التي يتم جمعها بالطرق التقليدية، وإن كانت عرضة لبعض التحفظات التقنية والأخلاقية.

وقد أورد الباحث عددا من الدراسات المختلفة، فضلا عن كيفية الاستعانة بالبيانات الضخمة في مجالات تخدم تطور الإنسان كالصحة والتعليم والاقتصاد، مؤكدا قدرة البيانات الضخمة على كشف المستور وتلاشي محاولات ادعاء المثالية والالتفاف على الاستبيانات المختصرة (دافيدوتس، 2018).

كما سبق أيضا للعديد من الباحثين محاولة استغلال البيانات الضخمة لفهم بعض الظواهر الاجتماعية، وقد كان لهذه الدراسات نتائج مذهلة يصعب الوصول إليها فعلا بالطرق الكلاسيكية لإجراء الدراسات الميدانية وجمع البيانات، ومن بين هذه الدراسات يمكن الإشارة إلى ما قام به "توبياس بريس" وآخرون حيث قاموا باستخدام بيانات اتجاهات جوجل (Google Trends) لإثبات أن مستخدمي الإنترنت من البلدان التي لديها ناتج محلي إجمالي أعلى للفرد (GDP) يتجهون للبحث عن معلومات حول المستقبل أكثر من المعلومات المتعلقة بالماضي. وتشير النتائج إلى أنه قد يكون هناك ارتباط بين السلوك عبر الإنترنت والمؤشرات الاقتصادية في العالم الحقيقي. وقد قام الباحثون في هذه الدراسة بفحص تسجيلات جوجل المصنوعة من قبل مستخدمي الإنترنت في 45 دولة مختلفة عام 2010، وقاموا بحساب نسبة حجم البحث للسنة التالية "2011" مقارنة بحجم البحث في السنة السابقة "2009" والذي أطلق عليه اسم "مؤشر التوجه المستقبلي". ثم قاموا بمقارنة التوجه المستقبلي مع الناتج المحلي الإجمالي للفرد الواحد في كل بلد، حيث وجدوا اتجاه قوي للبلدان التي يقوم فيها مستخدمي جوجل بالبحث عن المستقبل بالحصول على ناتج محلي إجمالي أعلى. وتشير النتائج لاحتمال وجود علاقة بين النجاح الاقتصادي للبلد وسلوك مواطنيها في البحث عن المعلومات المأسورة في البيانات الضخمة. (wikipedia)

8. تحديات استخدام البيانات الضخمة في البحوث الاجتماعية:

إن الاستفادة من البيانات الضخمة خاصة في الدول العربية يتطلب بداية وجود بنية تحتية جيدة واتصالا قويا بالإنترنت، إلى جانب امتلاك بعض الإمكانيات المعرفية التقنية الضرورية لمعالجة وتحليل البيانات والتحكم في لغات البرمجة المناسبة. تحدي الوصول إلى البيانات الضخمة: بالرغم من أن الكثير من البيانات المتاحة للعامة لديها استعمالات قوية في مختلف أهداف التنمية، إلا أن القطاع الخاص يحتفظ

بأحجام كبيرة منها، حيث أن الكثير من الشركات تمانع مشاركة البيانات خوفاً من المنافسة وحماية لزيائتها، ولهذا يتطلب العمل على البيانات الضخمة الشراكة بين منتجي البيانات ومستخدميها.

التحديات التحليلية وتحديات القدرات: إن عملية جعل البيانات الضخمة ذات معنى (لاستخراج المعلومات المناسبة) يتضمن مجموعة من المعضلات التحليلية التي يمكن أن تحد من صحة النتائج الناجمة عنها، حيث أن تحليل هذه البيانات يطرح الكثير من التحديات إما منهجية أو مرتبطة بصحة ترجمتها، أو المنهج المستخدم للتحليل من جهة (ابراهيمى، 2019، صفحة 52)، حيث أن التنوع الشديد لهذه البيانات وصعوبة فرزها وتنظيمها وإشكالية التأكد من صحة ومصداقية مصادرها وصعوبة ضبط وتتقية هذه البيانات... كل هذا يؤكد تحدي القدرة على استعمال كل القوة التي تتطلبها، ويدفع إلى ضرورة دعم وتقوية قدرات الباحثين على استخدام وتحليل البيانات الضخمة من جهة أخرى. وإلى جانب ذلك فإن التحدي الأساسي الذي يجب معالجته هو حماية خصوصية الأفراد وضمان السرية لبياناتهم وذلك لتفادي الاعتداء على حقوقهم والاستغلال السيء لبياناتهم في أغراض غير نزيهة. خاصة مع ضعف وعي الأشخاص بحجم وأنواع البيانات المولدة يوميا. ولهذا ينبغي العمل على توفير إطار قانوني وأخلاقي لمشاركة البيانات التي يتم الموافقة على جمعها واستعمالها بدون وعي وبدون فهم من أصحابها (ابراهيمى، 2019، صفحة 51). إذ أن نسبة كبيرة منها لم يؤخذ إذن أصحابها أو ليس لديهم علم أصلا بها، وحتى إن أبدوا موافقتهم على استخدامها فإن ذلك قد يكون دون وعي وإدراك كامل بذلك، أو تحت ضغط الحاجة لاستخدام بعض البرمجيات أو الوسائط التي تشترط الموافقة على سحب بيانات المستخدم.

وتعد هذه المسألة هي الإشكالية الأولى التي تحوم حول عملية استخدام البيانات الضخمة، فمن الضروري مراعاة الخصوصية والأخلاقيات في هذا الشأن، خاصة أن معظم البيانات تجمع بطريقة سلبية. وحتى إن تم حذف أو إخفاء البيانات الشخصية الصريحة فإن ذلك لا يضمن الحماية الكاملة لخصوصية الأفراد، إذ يمكن تحديد هوية الأفراد والجماعات من خلال جمع عدة مجموعات بيانات، وبالتالي يمكن تعريضهم للعديد من الأخطار الممكنة،

وكلما تنامي مقدار البيانات الشخصية والمعلومات الرقمية العالمية تزايد عدد الجهات التي تنفذ إلى هذه المعلومات وتستخدمها. فيجب تقديم تطمينات إلى أن البيانات الشخصية سئستعمل بصورة مناسبة، ومع التقيد بالقوانين ذات الصلة. فتتعيّن إعادة تقييم التهديدات والمخاطر المتأتية عن البيانات الضخمة، مع تكيف الحلول التقنية للتصدي لها.

9. خاتمة:

خلاصة القول أن تبني البيانات الضخمة في البحوث الاجتماعية قد أصبح أمرا ضروريا لا مناص منه، وهذا بالنظر لما يمكن أن يوفره من مزايا عديدة لا يمكن إنكارها، من أجل الوصول إلى نتائج أكثر كفاءة ودقة وأقل كلفة، غير أن هذا الأمر يتطلب دعما جادا للباحثين وللمؤسسات العلمية التي تسعى في هذا المجال، وذلك من خلال دعم القدرات المادية والتقنية والمعرفية لهذه المؤسسات التي تنسم في بلداننا العربية غالبا بالافتقار إلى البنية التحتية التكنولوجية والاقتصادية الكافية.

بالإضافة لندرة الموارد البشرية المؤهلة، وهذا أدى إلى إحداث نوع جديد من الفجوات التقنية وهي فجوة في تقصي البيانات الضخمة للوصول إلى نتائج دقيقة، مما يتطلب توفير برامج تكوينية للباحثين في المجالات الاجتماعية في استغلال البيانات الضخمة، حتى يتمكنوا من تحقيق أفضل النتائج الممكنة بالاستعانة بها بالكفاءة المطلوبة؛ إذ يحتاج الباحث في العلوم الاجتماعية إلى مزيد من الخبرة والتحكم في معالجة البيانات الضخمة من خلال الإلمام بأهم أساسيات علم الإحصاء، إلى جانب التحكم في إحدى لغات البرمجة الخاصة بمعالجة البيانات مثل لغة بايثون أو جافا، وكذلك الاطلاع على الآلية والذكاء الصناعي، وهي كلها ممكنة التعلم بسهولة للمبتدئين ولغير المتخصصين، حيث تتوفر العديد من الدورات التدريبية الموجهة لمختلف الفئات العمرية والعلمية لتعليم هذه التقنيات التي أصبحت لغة العصر وسمته، ويبدو أن من لا يتحكم فيها سيرج ضمن فئة الأميين مستقبلا.

كما يحتاج الباحث إلى قدرة واسعة على الإبداع وإعمال الخيال العلمي والسوسولوجي، حتى يتمكن من إيجاد روابط بين مختلف المؤشرات والمتغيرات. والقدرة على البحث عن المعلومات الصحيحة والدقيقة والمعبرة وسط الزخم الكبير من المعلومات الناقصة والخاطئة المزيفة أو غير الحقيقية.

إضافة إلى القدرة على الالتزام بأخلاقيات البحث العلمي وعدم الانجراف خلف الأضواء وحب الظهور والشهرة. مع ضرورة مراجعة الأطر القانونية من أجل تكريس الضوابط الأخلاقية الضرورية لحماية الأفراد والحفاظ على حقوقهم ومراعاة خصوصياتهم حيث أن البيانات الضخمة في كثير من الأحيان قد تجمع بانتهاكات للحياة الشخصية وقد يساء استخدامها.

10. قائمة المراجع:

Bibliographie

- (s.d.). Consulté le 04 25, 2022, sur www.marefa.org: www.marefa.org
- (s.d.). Consulté le 03 15, 2022, sur www.wikipedia.org
- (s.d.). Consulté le 03 05, 2022, sur www.un.org/ar/global-issues/big-data-for-sustainable-development
- Allemand, S. (2018). *Introduction aux technologies et applications Big. ANF.*
- BAMBA, K. (s.d.). *Comprendre le BIG DATA.*
- ITU. (2013). *Big Data today: normal tomorrow! ITU Technology Watch Report.*
- Laney, D. a. (2013). *Big Data Means Big Business. Gartner Inc.*
- Miele, S. a. (2013). *Analytics: The real-world use of big data. IBM Global Services.*
- un.org. (s.d.). *big-data-for-sustainable- development. Consulté le 03 15, 2022, sur www.un.org/ar/global-issues/big-data-for-sustainable-development*
- United Nations Global Pulse. (2013). *Big Data for Development A Primer.*
- wikipedia. (s.d.). *Récupéré sur www.wikipedia.org*
- ابراهيمى، أ. ص. (2019، 12 31). استغلال البيانات الضخمة لأغراض التنمية المستدامة في الدول العربية - الفرص والتحديات -. مجلة أرساد للدراسات الاقتصادية والإدارية. 49، (2)، p. 2،
- شيكور، ع. ا. (2019). استخدامات البيانات الضخمة في نظم المعلومات التسويقية. مجلة أبحاث كمية ونوعية في العلوم الاقتصادية والإدارية. 77، (2)، p. 1،
- عبدالسلام، م. (2021). تقنية البيانات الضخمة. أبوضبي: صندوق النقد العربي.

- لطايبي م, (2019, 1 2). البيانات الضخمة وصناعة المعلومات. مجلة الحكمة للدراسات الإعلامية والاتصالية. 6(4), p. 60.
- مركز الإحصاء). د.ت. (مفاهيم عامة حول البيانات الكبيرة، أدلة المنهجية والجودة، دليل رقم (13). أبوطيبي: مركز الإحصاء.
- . (s.d.). Consulté le 04 20, 2022, sur منصة المستقبل العلمي <https://www.platform4all.info/threads/84/>
- موسوعة عريق: (s.d.). Consulté le 04 20, 2022, sur <https://areq.net/m/%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA.html>
- . (s.d.). Consulté le 04 18, 2020, sur <https://areq.net/> موسوعة عريق/
- . (s.d.). Récupéré sur www.marifa.org موسوعة معرفة
- . (s.d.). Consulté le 04 18, 2022, sur نشرة الاتحاد المصري للتأمين https://www.ifegypt.org/NewsDetails.aspx?Page_ID=1244&PageDetailID=1386