البحث التطبيقي

جمهورية مصر العربية

معهد التخطيط القومي

الدراسات العليا

عنوان البحث باللغة العربية	البيانات الضخمة ودورها في تحسين جودة المؤشرات الاجتماعية
عنوان البحث باللغة الإنجليزية	Roles of big data in improving social indicators quality
إعداد	احمد جمال محمد رمضان
إشراف	د. أن تيسير احمد نصير

لاستكمال متطلبات

الحصول على الماجستير المهني "التخطيط للتنمية المستدامة"

عام 2022

الملخص

شهد العالم في السنوات الماضية انفجارا في كمية البيانات الرقمية المتاحة عبر الأقمار الصناعية ومختلف قنوات التواصل. ونظرا لثورة الرقمنة التي نشطت بفعل أجهزة ذات كفاءة عالية وسعر منخفض، ارتفع حجم البيانات المتاحة عبر العالم، وبعد اعتماد الجمعية العامة للأمم المتحدة في سبتمبر 2015 لأهداف التنمية المستدامة، أصبح من الضروري على الأجهزة الإحصائية الوطنية القيام بثورة في عملية جمع البيانات وطرق تحليلها في إطار رصد أهداف التنمية المستدامة بجميع أبعادها الاقتصادية والاجتماعية والبيئية[1].

أضافت أزمة جائحه كورونا عائق جديد في تنفيذ الإحصاءات الوطنية عن طريق الطرق الروتينية التي يعرضها اما للتأجيل أو الإلغاء، لذا كان من الضروري البحث عن مصادر غير اعتيادية لإنتاج الإحصاءات الرسمية بجودة ومصداقية تتقارب مع الأساليب التقليدية.

نستعرض في هذه الدراسة التطبيقية بعض مصادر البيانات الضخمة وانسبها التي يمكن الاستفادة منها في إنتاج الإحصاءات والمؤشرات الوطنية، ولتحقيق هذا الهدف تم صياغة مجموعه من الفروض أهمها "لا توجد علاقة ذات دلاله إحصائية بين استخدام تحليل البيانات الضخمة وتطوير إنتاج المؤشرات الاجتماعية"، " لا توجد علاقة ذات دلالة احصائية بين استخدام البيانات الضخمة وتقليل فجوات البيانات الاجتماعية". ولاختبار هذه الفروض تم الاستعانة بمجموعة من الأساليب الإحصائية لخدمة إعداد تلك الدراسة.

يقدم هذا التقرير نبذة عن المفاهيم المختلفة للبيانات الضخمة والتقصي عن أكثر تلك المفاهيم ملائمه للوضع الوطني ويستعرض بعد الجهود الدولية لتعظيم الاستفادة منها، كما يرصد بعض الخبرات الإقليمية حول استخدام البيانات الضخمة في تحسين جودة الإحصاءات الرسمية في إنتاج المؤشرات الاجتماعية.

يرصد التقرير نتائج التحليل الرباعي للنظام الإحصائي الوطني حول إمكانيه استخدام تحليل البيانات الضخمة في تحسين جودة إنتاج المؤشرات الاجتماعية، كما يعرض بعض الجهود الوطنية نحو مواجه تلك العقبات وتعظيم الاستفادة منها في المجالات الإحصائية المختلفة.

ويخلص التقرير الى ان النظام الإحصائي الوطني يستطيع خلال لفترة الحالية الاعتماد بشكل كبير على تحليل البيانات الضخمة لسد فجوات البيانات ولا سيما مؤشرات التنمية المستدامة لرفع جودة الإحصاءات الرسمية بشكل عام وإنتاج المؤشرات الاجتماعية بشكل خاص.

الكلمات المفتاحية: تحليل البيانات الضخمة، المؤشرات الاجتماعية، التحليل الرباعي

1- الأمم المتحدة. البيانات الضخمة وتحديث النظم الإحصائية

Abstract:

In the past years, the world has witnessed an explosion in the amount of digital data available via satellites and various communication channels. Due to the digitization revolution that has been stimulated by high-efficiency devices at low price, the volume of data available across the world has increased, and after the adoption of the Sustainable Development Goals by the United Nations General Assembly in September 2015 it became necessary for national statistical agencies to revolutionize the process of data collection and methods of analysis within the framework of Monitoring sustainable development goals in all its economic, social and environmental dimensions1.

The Corona pandemic crisis has added a new obstacle in the implementation of national statistics through the routine methods that it presents either for postponement or cancellation, so it was necessary to search for unusual sources to produce official statistics with a quality and credibility converging with traditional methods.

In this applied study, we review some sources of big data and the most appropriate ones that can be used in the production of national statistics and indicators. There is no statistically significant relationship between the use of big data and the reduction of social data gaps. To test these hypotheses, a set of statistical methods were used to serve the preparation of this study.

This report presents an overview of the different concepts of big data and investigates which concepts are most appropriate to the national situation and reviews international efforts to maximize the use of it. It also monitors some regional experiences on the use of big data to improve the quality of official statistics in the production of social indicators.

The report monitors the results of the four-year analysis of the national statistical system on the possibility of using big data analysis to improve the quality of production of social indicators. It also presents some national efforts towards facing these obstacles and maximizing the use of them in various statistical fields.

The report concludes that the national statistical system can, during the current period, rely heavily on big data analysis to bridge data gaps, especially sustainable development indicators, to raise the quality of official statistics in general and to produce social indicators.

Keywords: big data analysis, quadrant analysis, Social Indicators

المقدمة

تعد البيانات أساس التخطط الدقيق لتحقيق التنمية الشاملة ، حيث ترصد واقع المجتمع بشكل دقيق من خلال وصف احتياجاته ورغباته في شتى المجالات، كما توفر البيانات رؤيه مبدئية حول فرص التنمية المتاحة وفقا للخصائص الديموغرافية للسكان .لذا فإن توفير البيانات الدقيقة المحدثة بشكل مستدام أحد أهم اولويات التنمية الشاملة لتحقيق رفاهيه الانسان، وفى هذا الإطار ظهرت العديد من المبادرات الدولية نحو رصد وتقيم الأحوال المعيشية للإنسان بشكل عام وإعداد العديد من المؤشرات لرصد ذلك بدأيه من قياس الموارد البشرية بالإضافة الى عدالة التوزيع في الدخل مرورا بمبادرة الأمم المتحدة في اطلاق الأهداف الإنمائية للألفية وصولا الى أهداف التنمية المستدامة بعنوان لا احد يتخلف عن الركب والتي تنتهى بنهاية عام 2030.

ومع تعاقب السنوات أصبحت المؤسسات البحثية والجهات الحكومية المختلفة وعلى راسها الاجهزة الإحصائية تمتلك ثروه هائلة من البيانات المتنوعة في شتى المجالات من خلال الدراسات الميدانية والبحوث الإحصائية التي تقوم بها تلك الهيئات بشكل مستمر والتي يمكن الاستفادة منها معا في إطار تقنى جيد يراعى فيه المفاهيم الإحصائية ويلبى الطلب المستمر والتغير على البيانات والذى يختلف باختلاف الوقت والتوجه العام للسياسات المحلية والدولية، لذا كان من الضروري ان يتم تطوير شكل البيانات ومحاوله الاستفادة من رقمنه الحياة البشرية في شتى المجالات.

ومع ظهور جائحه كورونا وما ترتب عليها من اجراءات وقائية للحد من انتشار الفيروس والذي بدوره حد من تنفيذ العمليات الإحصائية الدورية التي تقوم بها الأجهزة بالطرق المعتادة تنفيذا للإجراءات الاحترازية لمواجهة تفشى الفيروس، مما ادى الى حدوث فجوات في توفير البيانات، مما دعي الاجهزة الإحصائية في البحث على مصادر غير تقليدية للبيانات والاستفادة من البيانات المتوفرة لديها باستخدام تقنيات البيانات الضخمة(big-data) لسد فجوات البيانات والحصول على رؤية مقربه للواقع لرصد خصائص المجتمع والتنبؤ بالوضع في المستقبل.

بعد إطلاق الرؤية الوطنية مصر 2030 والتي استهدفت بالأساس رفض مستوى معيشية المواطن المصري وتلبية احتياجات الاساسية، لذا كان من الضروري الاعتماد على مجموعه من مؤشرات الأداء التي ترصد مستوى معيشية المواطن تستند الى مصادر بيانات جيدة ذات فترات زمنية متقاربه، ولكن نجد ان هناك العديد من الفجوات في المؤشرات الاجتماعية منها الفترات متباعدة في السلاسل الزمنية من 2 الى 10 سنوات، كذلك نجد ان بعض تلك المؤشرات لا يتم أصدرها بصفه دورية نظرا لوجود بعض الصعوبات في التمويل.

وتعمل تقنيات البيانات الضخمة على تحليل البيانات من مصادر مختلفة مستقله، حيث ان تحليل هذه البيانات يسمح بوجود ارتباطات بين مجموعة من البيانات المستقلة لكشف جوانب عديدة، وتتعدد استخدامات البيانات الضخمة في كثر من المجالات ومنها توفير ادوات مبتكرة لصانعي القرارات لفهم أفضل للظروف مما يساعد في اتخاذ القرارات الصحيحة التي تحقق الاهداف المنشودة، ولكن تبرز في سياق متصل تظهر مجموعه من العقبات الفنية والتقنية ولاسيما الأخلاقية في استخدام تلك الطرق في التحليل وإنتاج المؤشرات الوطنية ومدى توافقها مع مبادئ الإحصاءات الرسمية، والى أي مدى يمكن الاعتماد على نتاج تلك الطرق في إعداد خطط التنمية ورصد مدى تنفيذها.

المشكلة البحثية

في خضم التطور العلمي والثورة التكنولوجية الهائلة التي يشهدها العالم في العقود الأخيرة وما واكبها من زخم في كمي ومعرفي في شتى المجالات ، أخذت البيانات الضخمة اهتماما متزايد بمصادرها وطرق تحليله وكيفية تعظيم الاستفادة منها ، وفي هذا الشأن فقد فرضت طبيعة تلك البيانات الضخمة تحديثات كبيرة على طبيعة أعمال المنظمات في مختلف المجالات (Kitchin,2014)،حيث يمكن القول أن البيانات أصبحت مصدر قوة رئيسي لأى مجتمع قائم على المعرفة ، حيث البيانات الضخمة في حال إدارتها بشكل سليم وموضوعي تسهم إسهاما مؤثرا في التنمية الاقتصادية والاجتماعية.

في اطار تنفيذ خطط التنمية تبرز الحاجه الملحة لتوفير بيانات ومؤشرات حقيقيه بشكل مستمر يعكس الوضع الاقتصادي والاجتماعي بطريقه مرونة دون اضافه اعباء اقتصادية أخرى، مما يتطلب توفير مصادر غير تقليدية للبيانات لإنتاج مؤشرات تعمل على رصد وتقييم المجتمع بخصائصه المختلفة، حيث نجد أن التقرير الأخير لرصد أهداف التنمية المستدامة يغطى فقط (48%) من المؤشرات مما يوضح حجم فجوة البيانات ، ويبرز هنا تحليل البيانات الضخمة والتي تعتمد بشكل أساسي على استخدام البيانات المتاحة التي تم تجميعها سابقا ودمجها مع العديد من البيانات مفتوحه المصدر بإجراء التحليل المناسبة للاستفادة من تلك البيانات.

وهنا يمكن التساؤل الرئيسي في البحث " هل يمكن الاعتماد على البيانات الضخمة في تحسين جودة إنتاج المؤشرات الاجتماعية بطريقة تتناسب مع التعاريف الدولية وتوفر الرؤية الصحيحة لصانعي القرار في إعداد خطط التنمية" وينبثق من هذا السؤال مجموعه من الأسئلة الفرعية:

- ما المقصود بمفهوم البيانات الضخمة وما هي أهم خصائصها؟
- هل هناك تحديات تواجه البيانات الضخمة وتحليلها؟
- هل يمكن الاعتماد على البيانات الضخمة وتحليلاتها في إنتاج المؤشرات الاجتماعية؟
- هل يمكن استخدام تحليل البيانات الضخمة في إعداد خطط التنمية؟
- هل تمتلك الأجهزة الإحصائية القدرة والرغبة نحو استغلال كنوز البيانات في إطار إحصائي تقني سليم؟
- هل هناك تجارب ناجحة في دول مشابهه لاستخدام البيانات الضخمة، وما مدى الاستفادة منها؟

أهداف البحث

يسعى الباحث في هذا الدراسة الى تحقيق مجموعه من الأهداف منها:

1- التعرف على المفاهيم المختلفة للبيانات الضخمة وانواعه.

2- رصد الفرص والتحديات نحو استخدام تحليل البيانات الضخمة في تحسين جودة إنتاج المؤشرات الاجتماعية.

3- رصد تجارب الدول المشابه اجتماعيا في استخدامات تحليل البيانات الضخمة.

4- رصد فرص استخدام تحليل البيانات الضخمة في إعداد خطط التنمية ورصد أهداف التنمية المستدامة.

أهمية البحث

تنبع أهمية الدراسة من الناحية الأكاديمية والعملية فيما يلي:

- الاهتمام الوطني بالحصول على مؤشرات دقيقه لرصد الأحوال المعيشية للمواطن بشكل دوري مستدام.
- إيجاد بدل فني جيد للحصول على البيانات في حالات الاغلاق الكلى والجزئي بسبب الازمات المختلفة.
- اهتمام كافة العديد من الجهات المحلية والإقليمية والدولية بتحليل البيانات الضخمة وتأثيراتها المختلفة.
- البحث عن مصادر مبتكرة لسد فجوات البيانات المفقودة الاجتماعية، بالأخص في مؤشرات التنمية المستدامة
- ندرة الدراسات الوطنية في مجالات الاستفادة من تحليل البيانات الضخمة خصوصا في المجالات الإحصائية.

حدود البحث

- الحدود الموضوعية: تسعى الدراسة الى دارسة الاستفادة من البيانات الضخمة في رفع جودة وكفاية العمل الإحصائي في إنتاج المؤشرات الاجتماعية بشكل عام ومؤشرات قياس مستوى المعيشة بشكل خاص.
- الحدود المكانية: الجهاز المركزي للتعبئة العامة والإحصاء– وزارة التخطيط–جامعة القاهرة.
- الحدود البشرية: العاملين بقطاع الإحصاءات السكانية بجهاز الإحصاء الوطني – مجموعة من مستخدمي البيانات بوزارة التخطيط – مجموعه من الأكاديميين بكلية سياسية واقتصاد –جامعه القاهرة.
- الحدود الزمنية: تم إجراء البحث في عام 2022.

فرضيات الدراسة

تسعى الدراسة بشكل أساسي على دارسة إمكانية استخدام تحليل البيانات الضخمة في دعم جودة الإحصاءات والمؤشرات الاجتماعية، وذلك من خلال مجموعه من الفرضيات والتي يمكن اجمالها في الفرضيات التالية:

- لا توجد علاقة ذات دلالة احصائية بين أراء المستجيبين حول أهمية تحليل البيانات الضخمة.
- لا توجد علاقة ذات دلاله إحصائية بين استخدام تحليل البيانات الضخمة وتطوير إنتاج المؤشرات الاجتماعية.
- لا توجد علاقة ذات دلالة احصائية بين استخدام البيانات الضخمة وتقليل فجوات البيانات الاجتماعية.
- لا توجد علاقة ذات دلالة احصائية بين استخدام تحليل البيانات الضخمة واعدد خطط التنمية الاجتماعية.

منهجية البحث

توافقا مع طبيعة المشكلة البحثية محل الدراسة وسعيا نحو الوصول الى إجابات أكثر دقه وموضوعيه للأسئلة البحثية، فقد تم الاعتماد على المنهج الوصفي التحليلي وفى وصف وتفسير الظاهرة محل البحث بأبعادها المختلفة، وتم الاعتماد على مجموعه من المصادر والادبيات ذات الصلة بموضوع البحث تتم بالحداثة والموضوعية كأداة لجمع المعلومات ومعالجتها للوصول الى أهداف البحث، بالإضافة الى القيام بدراسة تطبيقيه لخدمه أغراض البحث واختبار فروضه الإحصائية لتحقيق هدف الدراسة.

الدراسات السابقة

أولا: الدراسات العربية

(الأمم المتحدة،2014) بعنوان "البيانات الضخمة وتحديث النظم الإحصائية":

تقرير أعدته اللجنة الإحصائية بالأمم المتحدة بهدف تقييم استخدام البيانات الضخمة، في الوقت الراهن، لأغراض الإحصاءات الرسمية. ويقدم التقرير استعراضا عاما للأنشطة التي اضطلعت بها الدوائر الإحصائية الرسمية من أجل تطبيق استخدام البيانات الضخمة في الإحصاءات الرسمية. وأبرز التقرير الأهمية المتنامية للبيانات الضخمة في شتى المجالات بما في ذلك الإحصاءات الرسمية وضرورة الاعتماد على البيانات الضخمة في زمن قل فيها الاعتماد على المصادر التقليدية لجمع البيانات، واستعرض التقرير بعض العقبات التي تواجه استخدام البيانات الضخمة والتي تركزت في مشكلتين أساسيتين وهما: تركز مصادر البيانات الضخمة في القطاع الخاص فقط، ثانيا عدم وجود تشريعات تسمح باستخدام البيانات الضخمة لأغراض إحصائية.

وسرد التقرير الإطار المفاهيم للبيانات الضخمة من أنواعها واشكالها بشكل متوافق عليه دوليا وعرض بعض المجهودات الدولية في إطار التوجه نحو استخدام البيانات الضخمة في الإحصاءات الرسمية، كما أبرز التقرير مجموعه من التوصيات الوارد من الأجهزة الإحصائية بدول الاتحاد الأوروبي نحو كيفية التحول لاستخدام البيانات الضخمة في إنتاج الإحصاءات الرسمية وابداء الرغبة في نقل الخبرات للأجهزة الإحصائية الأخرى.

كما اظهر التقرير الى الحاجة إلى إطار يحدد القواعد الأساسية لسبل التوليف بين مختلف مجموعات البيانات الضخمة وحمايتها وتبادلها وعرضها وتحليلها والاحتفاظ بها. فبوجود هذا الإطار، تنتهي المشكلة الكبيرة المتعلقة باكتساب اطمئنان عموم الناس إلى سلامة استخدام الحكومة لبياناتهم الشخصية، حيث تجاوب الناس تجاه استخدام البيانات الضخمة لأغراض الإحصاءات الرسمية عاملا حاسما في كفالة أي استخدام لهذه البيانات في المستقبل. فاكتساب ثقتهم أمر ملح مطلوب لضمان استخدامها بالطريقة المثلى.

واقترح التقرير طريقا لتطوير استخدام البيانات الضخمة في الأغراض الإحصائية وذلك بإنشاء فريق عامل على الصعيد العالمي، يستند إلى مبادرات إقليمية قائمة تسعى إلى مشاطرة الفرص في مجالات التطورات المنهجية، وأفضل الممارسات المتصلة بالمسائل الاستراتيجية، كما يوفر فرص التدريب المناسبة لبناء القدرات الوطنية في مجالات استخدام البيانات الضخمة لإنتاج الإحصاءات الرسمية بكافه أشكالها ، كما سيقوم الفريق أيضا بتيسير الشراكة الدولية لاستخدام البيانات الضخمة في نقل التكنولوجيا إلى البلدان النامية وفي دعم خطة التنمية.

(ركتن، 2018) بعنوان " البيانات الضخمة ما أهميتها وما أهمية الاستثمار في تحليلها، وكيف ستؤثر في حياتنا وقرارتنا"

يتناول التقرير بعض الاستخدامات ومزايا تطبيق البيانات الضخمة على المستويات الوطنية والمؤسسية فيشير التقرير إلى أن البيانات الضخمة أصبحت تستعمل كأداة للدراسات وبناء الاستراتيجيات والتقييم من جهة ،وتحدي للمنظمات الدولية ، وغيرها من المنظمات من جهة أخري، حيث تعمل تطبيقات البيانات الضخمة على تعزيز الكفاءة

التشغيلية لتلك المنظمات ، واتخاذ القرار المناسب وتحقيق قيمة مضافة ، وذلك من خلال التوصيف الدقيق والصحيح لماهية البيانات الضخمة ، وكيفية الاستفادة من تقنياتها وتحديد أفضل الممارسات للمنظمات وسالمة بياناتها . وفي هذا الصدد فقد أشارت تلك الدراسة ان البيانات الضخمة تساهم بشكل كبير في بناء العديد من المجالات في كثير من الدول النامية، كما إن التركز الجغرافي لعلماء البيانات في بعض البلدان كأمريكا قد يزيد من ظاهرة الهجرة من الدول العربية نحو هذه البلدان وليس فقط تصدير البيانات أما عن الأثر غير المباشر للبيانات الضخمة على الاقتصاد، فإن مراقبة ومتابعة ومشاركة المعلومات في الوقت الحقيقي، في الكفاءة عبر مراحل الانتاج المختلفة.

وخلص التقرير ان البيانات الضخمة تساعد بشكل كبير في دعم عملية انتاج البيانات والمؤشرات بشكل عام خاصه في رصد الخصائص الاجتماعية للسكان، ولكن مع ضرورة إجراء المراجعات الإحصائية اللازمة لاعتماد تلك النتائج وتحديد مستوى تغطيتها الجغرافية للسكان.

(ملياني فتيحة، 2019) بعنوان "البيانات الضخمة: الفرص، التحديات، ومجالات التطبيق"

تهدف الدراسة الى وضع إطار مفاهيمي للبيانات الضخمة من خلال تحديد مفهومها، تقنيا، خصائصها، مصادرها ثم توضيح الأطراف في منظومة البيانات الضخمة ، كما تقوم برصد مجالات استخدامات البيانات الضخمة في التجارب العالمية الناجحة، يأتي في مقدمة تلك المجالات إنتاج المؤشرات الإحصائية المختلفة بجودة توازى التي يتم إنتاجها بالطرق التقليدية ، كما يرصد إمكانية الاستفادة منها عربيا وذلك برصد الفرص التي تساعد على التوجه نحو استخدام البيانات ويأتي في مقدمتها أن تحليل البيانات الضخمة يساعد في التوصل إلى فهم أكثر شمولا وتفصيلا لخصائص واحتياجات الأفراد والمجموعات المختلفة، وذلك عن طريق إدراج أجزاء متعددة ومتباينة من البيانات في عملية التحليل، وبالتالي يمكن أن تكون الخدمات والعروض المقدمة لهذه الجماعات أو الأفراد أكثر تخصيصا وملاءمة ..وغيرها من تلك الفرص وكما تطرقت الدراسة الى بعض التحديات والتي يأتي في مقدمتها طبيعة البيانات الغير منتظمة حيث ان أكثر من 80% من تلك البيانات غير منظمة مما يمثل تحدى كبير في تحويلها للشكل المناسب الذى يتيح استخدمها بالإضافة الى صعوبة تحقيق مبدا خصوصيه البيانات مع تضاعف حجمها وسرعه تبادلها وغيرها من التحديات ، كما اكدت الدراسة الميزة التنافسية لاستخدام البيانات الضخمة في المجالات الإحصائية لما له دور في زيادة جودة ودقه العمل الإحصائي ، تم الاعتماد على المنهج الوصفي في تحليل موضوع الدارسة.

وخلصت الورقة البحثية الى أن هناك العديد من المجلات الأخرى التي تساعد البيانات الضخمة في تطويرها وتمكينها من توفير خدمات أفضل في المستقبل، كما انها توفير اليه مناسبة لإنتاج مؤشرات إحصائية انية تعمل على رصد الواقع ومتابعه تنفيذ برامج التنمية بطريقه أكثر تفاعلية.

(مقناني 2019) بعنوان "دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية":

تهدف هذه الدراسة الى رصد أوجه استخدام البيانات الضخمة في دعم تحقيق أهداف التنمية المستدامة وتسخير البيانات في صنع القرار من خلال بناء أنظمة البيانات المبدعة التي تعتمد على مصادر البيانات المحدثة، كما

تسعى الى توفر معلومات عن البيانات الضخمة، وتهدف إلى تبيان أن البيانات مصدر قوة ومن شأنها أن تسهم بشكل جذري في التنمية المستدامة، وتؤثر في النهوض بالمجتمع العربي خاصة وأننا في ظل مجتمع المعرفة.

تم إعداد تلك الدراسة بالاعتماد على المنهج الوصفي التحليلي في وصف وتفسير الظاهرة وأبعادها .

وابرزت الدراسة في نتائجها الى تعدد المجالات التي يمكن استخدام البيانات الضخمة بها ويأتي في مقدمتها الأغراض الإحصائية التي تعمل على متابعة أهداف التنمية المستدامة بما في ذلك البعد الاجتماعي، كما دعت الدراسة الى ضرورة تبيني الحكومات العربية البيانات الضخمة بمؤسستها المختلفة في استراتيجيتها الإنمائية، كما ابرزت إن تنفيذ السياسات لإطار تأمين المعلومات الوطنية هو عامل جوهري لاعتماد البيانات الضخمة من أجل تحفيز الابتكار، والتطور، والرقي، إذ توفر البيانات الضخمة القدرة نطاقا غير مسبوق لفهم المجتمع بخصائصه. كما أوصى التقرير بضرورة اعتماد التعامل مع البيانات الضخمة في إطار تشريعي متكامل بين كافة الجهات المنتجة والمستخدمة للبيانات لضمان جودة ودقه تلك البيانات للاعتماد عليه.

(الاسكوا،2020)بعنوان "استخدام البيانات الضخمة في احتساب مؤشرات أهداف التنمية المستدامة في المنطقة العربية التحديات والفرص":

تقرير فني اصدرته منظمة الإسكوا يهدف الى تحديد نطاق مفهوم البيانات الضخمة في الإحصاءات وتحديد إيجابيات وسلبيات ومزايا وعيوب استخدام البيانات الضخمة للإحصاءات الرسمية بشكل عام وفى إنتاج الإحصاءات الرسمية بشكل خاص، وذلك في محاولة لمعرفة الدور الذي يمكن أن تلعبه البيانات الضخمة للإحصاءات في المستقبل القريب. كذلك يقوم بمراجعة للمشاريع الفعلية الجارية في البيانات الضخمة للإحصاءات الرسمية وعرض أمثلة على استخدام مصادر البيانات الضخمة للاستشعار عن بعد لرسم خرائط مؤشرات الفقر، كذلك فانه يلقى نظرة عامة عن مصادر البيانات الضخمة ذات التطبيقات الملموسة، بالإضافة الى أنه يقوم بمناقشة التحديات والفرص المتعلقة بالبيانات الضخمة في تجميع أهداف التنمية المستدامة خصوصا في المنطقة العربية، حيث ابرز التقرير العديد من المزايا نحو التوجه لاستخدام البيانات الضخمة خصوصا في الاجهزة الإحصائية للدول متوسطة النمو من اجل توفير بيانات اكثر دقه باقل تكلفه بالإضافة الى تحسين التوقيت والأنية وامكانية التعمق في التفاصيل مع القابلية للمقارنات الدولية.

أستهدف التقرير جميع الدول العربية من خلال استبيان تم توزيعه على جميع الأجهزة الإحصائية الوطنية بعنوان "تكامل مصادر البيانات الغير تقليدية في احتساب مؤشرات أهداف التنمية المستدامة في بلدان عربية"

وخلص التقرير الى ان البيانات الضخمة يمكن الاعتماد عليها في إنتاج الإحصاءات الرسمية، ولكن هناك بعض العقبات امام تحقيق ذلك تتمثل في تحديات اخلاقية وفنية وقانونية وتحديات تتعلق بالسمعة، كذلك فالبيانات الضخمة تأتى غالبا من مصادر مفتوحة ولا تعطى الأولية الكافية لعمليه سرية وتامين البيانات.

ثانيا: الدراسات الأجنبية

(Journal of Official Statistics,2018) بعنوان

"Small Area Model-Based Estimators Using Big Data Sources"

تهدف الدراسة التطبيقية الى إعداد دليل عملي على إمكانية استخدام البيانات الضخمة بشكل أكثر فاعلية في إنتاج أكثر المؤشرات الاجتماعية أهمية وهو الفقر وذلك من خلال النهج متعدد التخصصات، الذي يجمع بين مجموعة الأبحاث الإحصائية في تقدير المساحة الصغيرة مع مجموعة البحث في التنقيب عن البيانات الاجتماعية بناءً على البيانات الكبيرة، والذي بدوره يمكن أن يوفر وسائل جديدة لمعالجة مشكلة المراقبة الدقيقة في الوقت المناسب لمؤشرات الفقر أو عدم المساواة، على نطاق مكاني وزمني دقيق.

تقترح هذه الدراسة ثلاث طرق لاستخدام البيانات الضخمة جنبًا إلى جنب مع تقنيات تقدير المنطقة الصغيرة، وتوضح كيف يمكن للبيانات الضخمة أن تعكس جوانب الرفاهية والظواهر الاجتماعية والاقتصادية الأخرى، حيث اعتمدت الدراسة على استخدام ثلاث مصادر للبيانات الضخمة وهي الشبكات الاجتماعية بمختلف اشكالها، الرسائل الإلكترونية وجهات الاتصال، سجلات مكالمات التليفون ومسار تحديد الموقع العالمي.

خلصت الدراسة الى ان البيانات الضخمة توفر مجهرًا قويًا يمكن أن يساعدنا في فهم العديد من الظواهر الاجتماعية والاقتصادية المعقدة والتنبؤ بها، من نشر المعلومات والابتكار والأزمات إلى التوزيع غير المتكافئ للموارد والفرص، خصوصا مجالات دراسة الفقر ومستوى المعيشية. بالإضافة الى أن البيانات الضخمة أسرع بكثير في توفير المتغيرات المساعدة لنماذج المناطق الصغيرة من مصادر البيانات الرسمية. كما تتميز البيانات الضخمة بالمرونة، ففي حالة عدم توفر البيانات الفردية بدى الخصوصية يمكن الاستفادة بالبيانات على مستوى المناطق كمتغيرات ترجيحية في إعداد نماذج تقدير معدلات الفقر.

(Sustainability 2019) بعنوان

"Big Data and Their Social Impact"

تهدف الدراسة هو استكشاف مدى دقة البيانات الضخمة والخدمات المستندة إلى البيانات الضخمة في التأثير على الأفراد والمجتمعات حيث تتهتم الدراسة بجوانب الخصوصية وسريه البيانات، حيث تتناول بالتفصيل تصورات الأفراد للبيانات، لا سيما حول كيفية إدراكهم للمشاركة الفعلية لبياناتهم، حيث تحدد هذه الدراسة مساحة قيمة للتأثير الاجتماعي للبيانات الضخمة ذات الصلة بثلاثة عوامل، وهي النية لمشاركة البيانات الشخصية، ومخاوف الأفراد، والتأثير الاجتماعي للبيانات الضخمة.

عملت الدراسة على إعداد إطار أولى لفهم التأثير الاجتماعي لبحوث البيانات الضخمة من خلال 3 محاور وهي مخاوف المستخدم، نية المستخدم لتبادل البيانات، التأثير الاجتماعي للبيانات الضخمة. يهدف هذا الإطار الى أن المستخدمين بطريقة ما، بتصوراتهم ونواياهم لاستخدام تطبيقات البيانات الضخمة، يحددون مساحة القيمة الشخصية الخاصة بهم وربما أيضًا مساحة القيمة المجتمعية.

انتهجت الدراسة التحيل الوصفي والكمي للبيانات من خلال دراسة شملت 108 مستجيب (18 –70) سنه.

خصلت الدراسة إلى إعادة النظر في الرؤى المتعلقة باستخدام أبحاث البيانات الضخمة من الأفراد، حيث ان الدافع الرئيسي لديهم هو لحماية خصوصيتهم وملكية بياناتهم الشخصية مع علمهم أن يمكن استخدامها في خدمات فريدة

تتحسن جودة حياتهم، كذلك اتفاق المشاركين على القيمة المضافة للبيانات الضخمة تجاه الخدمات الاجتماعية الواعية المستجيبة المشاكل الاجتماعية الحرجة.

(ECLAC،2020) بعنوان

"An assessment of big data for official statistics in the Caribbean"

تهدف هذه الدراسة الى دراسة دمج البيانات الضخمة في إنتاج الإحصاءات الرسمية في منطقة البحر الكاريبي من خلال مناقشة الفرص والتحديات التي تطرحها البيانات الضخمة بهدف تسليط الضوء على الفوائد المحتملة للبيانات الضخمة في الإحصاءات الرسمية، مع التعرف على العقبات التي تعترض ومخاطر دمج البيانات الضخمة في الإحصاءات الرسمية. كما تقدم أيضا اقتراحات حول كيفية التخفيف من هذه المخاطر واقتراح خطوات قابلة للتنفيذ لدمج البيانات الضخمة في الإحصاءات الرسمية في منطقة البحر الكاريبي.

يأتي التقرير في إطار سعى دول منظمه الكاريبي نحو البحث عن مصادر بيانات غير تقليدية لتوفير مصادر للبيانات وإنتاج مؤشرات التنمية المستدامة وذلك لوجود قصور شديد في الإحصاءات الوطنية حيث تصنف دول المنظمة بالدول الفقيرة في البيانات ليس فقط لأن البيانات محدودة ولكن أيضًا لأن البيانات المتاحة لا يتم نشرها على نطاق واسع وفي الوقت المناسب، وذلك نتيجة لمحدودية القدرات والموارد. كما عرض التقرير العديد من الجهود التي قامت بها منظمات الأمم المتحدة بهيئاتها المختلفة نحو توفير الدعم المناسب في إنتاج الإحصاءات الرسمية وتوفير وسائل التدريب المناسبة لرفع كفاءة الأجهزة الإحصائية في استخدام البيانات الضخمة.

عرض التقرير العديد من الفرص نحو استخدام البيانات الضخمة في دول منطقه الكاريبي يأتي في مقدمتها سرعه الحصول على النتائج وانخفاض التكاليف في إنتاج المؤشرات الوطنية بنسبة كبيرة، ومع الزيادة المستمرة في حجم البيانات أصبح من الضروري التوجه نحو استخدامها والاستفادة منها بالشكل المناسب. شارت الدراسة الى بعض التجارب الناجحة للاعتماد على مصادر البيانات الضخمة في التخطيط ومواجهة الازمات مثل ازمة تفشى الفيروسات كما حدث في ترينيداد وتوباغو 2014 وذلك بالاعتماد على بيانات المحمول، كما توفير البيانات الضخمة الفرصة الحقيقة لتحديد الفئات الأكثر احتياجا وتوفير التحذيرات الكافية لذلك بتكلفه أقل....

وخلصت الدراسة أن تعدد مزايا البيانات الضخمة ويأتي في مقدمتها أنه تؤثر بشكل إيجابي على التنمية من خلال العمل كنظام إنذار مبكر لتسهيل الاستجابات السريعة في الأزمات، كما دعت الى ضرورة البدء في مزيد من المناقشات حول العديد من جوانب البيانات الضخمة وتوفر منصة أولية يمكن بناء استراتيجية البيانات الضخمة في منطقة البحر الكاريبي، وذلك من خلال إطار مفاهيمي عملت الدراسة على إعداده بعنوان "نحو استراتيجية البيانات الضخمة لمنطقة البحر الكاريبي"

(ADB، 2020) بعنوان

“MAPPING POVERTY THROUGH DATA INTEGRATION AND ARTIFICIAL INTELLIGENCE”

تقرير صادر عن بنك التنمية اسيا بالتعاون مع جهاز الإحصاء الفلبيني من اجل استخدام خوارزميات الذكاء الاصطناعي والشبكات العصبية في تحليل البيانات الضخمة لرصد معدلات الفقر بدولة الفلبين. ويتعرض التقرير لأهمية استخدام البيانات لمواجه الفقر والسعي لاستخدام بيانات الاقمار الصناعية كأحد مصادر البيانات الضخمة لتقدير معدلات الفقر بالتوازي مع الإحصائيات الرسمية، حيث تم انشاء لجنه عام 2015 بالتعاون مع الشركاء المحليين للاستفادة من البيانات جغرافية وبيانات الأرض في رصد أهداف التنمية المستدامة.

تم الاستفادة من خرائط الفقر التي تنتجها كلا من الفلبين وتايلاند لتحديد توزيع الفقراء واستخدامها في تدريب النموذج المعد لتحليل البيانات الضخمة الخاصة بصور الأقمار الصناعية. ويعرض التقرير النتائج الأولية للدراسة التي تم إجراؤها من خلال التعاون بين بنك التنمية الآسيوي، والأجهزة الوطنية للإحصاء في الفلبين وتايلاند، ومختبر البيانات العالمي.

ودراسة امكانية اعتماد استراتيجيات بديلة لجمع البيانات من خلال الجمع بين طرق المسح التقليدية ومصادر البيانات المبتكرة لتعزيز دقة إحصاءات الفقر وحسن توقيتها وفعاليتها من حيث التكلفة وجودتها. ويعمل التقرير على المقارنة بين الطرق التقليدية لتقدير معدلات الفقر ومزايا تطبيق تحليل البيانات الضخمة لتوفير تلك المؤشرات. كما اعتمدت الدراسة على دمج البيانات الضخمة في نموذج تقدير المناطق الصغيرة كأحد المدخلات لبناء خريطة الفقر. واستعراض التقرير الطرق الأكثر شيوعا لتقدير معدلات الفقر باستخدام البيانات الجغرافية ومنها يتم بناء نموذج هيكلي للتنبؤ بمعدلات الفقر باستخدام البيانات المشتركة مع البيانات المكانية وغيرها من المعلومات التي يمكن اشتقاقها من صور الأقمار الصناعية مثل كثافة المباني وعددها بالإضافة الى عدد السيارات وكثافه الطرق ونوع التربة الزراعية ... وغيرها من المؤشرات واظهرت النتائج الارتباط الكبير بين المتغيرات التي تم استخلاصها من صور الأقمار الصناعية ومعدلات الفقر المتوقعة بنسب تصل الى 60%.اما الطريقة الاخرى فهي طريقه أكثر تجردا والتي تعتمد على نماذج الشبكات العصبية والتعليم الألى العميق حيث يتم التعامل مع كافه المتغيرات دون الحاجه الى تحديد مسبق لها ويقوم النموذج بتقدير معدلات الدخل وعلى اساسه يتم تقدير معدلات الفقر. وتتميز الطريقة الثانية بالجودة العالية في عمليات التنبؤ. كذلك استعرض التقرير العديد من الاساليب والدراسات في نفس السياق والتي تعتمد على تحليل بيانات الاضاءة الليلية لتقدير معدلات الفقر كذلك تقدير معدلات النمو الاقتصادي وغيرها من الاستخدامات الاقتصادية التي تعكس العديد من الخصائص الديموغرافية والاقتصادية للسكان في العديد من دول العالم. واستعرض التقرير عمليه انشاء طبقات الخلايا العصبية وكيفية تدريبها باستخدام بيانات خريطة الفقر واستخدام الصور الفضائية عالية الجودة لتقدير والتنبؤ بمعدلات الفقر.

وخلص التقرير الى التأكيد على اهميه تبنى البيانات الضخمة كمصدر أساسي للإحصاءات الرسمية وضرورة العمل على توفير الدعم الفني والتقني والتدريب للعاملين بالأجهزة الاحصائية نحو استخدام وتطوير تلك الاساليب في تحليل البيانات ودراسة الاثار الجانبية لاستخدام البيانات الضخمة على شركاء العمل الإحصائي من الأكاديميين والمنشآت البحثية وغيرها من المؤسسات ذات الصلة بالعمل الإحصائي وذلك لضمان منصة واسعة للمشاركة الأفكار والمعرفة والحلول حول كيفية الاستفادة مصادر البيانات المبتكرة لصالح الجميع.

(ESCAP، 2021) بعنوان

"Big data for population and social statistics"

يأتي هذا التقرير ضمن مجموعه من التقارير التي تصدرها الشعبة الإحصائية في اللجنة الاقتصادية والاجتماعية لآسيا والمحيط الهادئ في إطار استخدام مصادر البيانات غير التقليدية للإحصاءات الرسمية، ويهدف هذا التقرير الى نظرة عامة على مصادر البيانات الضخمة التي يمكن استخدامها لإنتاج الإحصاءات السكانية والاجتماعية وتقدم أمثلة في استخدام بيانات الهاتف المحمول وبيانات المبيعات وبيانات رصد الأرض ووسائل التواصل الاجتماعي وغيرها من البيانات عبر الإنترنت لإنتاج إحصاءات متعلقة بالنقل وإحصاءات العمل ، الإحصاءات الصحية ومؤشرات أهداف التنمية المستدامة.

أبرز التقرير المصادر المختلفة للبيانات الضخمة التي يمكن الاعتماد عليها لإنتاج المؤشرات السكانية والاجتماعية والتي يأتي في مقدمها السجلات الإدارية التي يمكن الاستفادة منها في اجراء التعدادات كذلك بيانات المحمول التي يمكن الاعتماد عليها في إحصاءات الهجرة بأنواعها والسياحة. كما تطرق التقرير الى إمكانية الاعتماد على الوظائف المتاحة اونلاين وذلك للاستخدام في إحصاءات العمالة. كذلك أبرز التقرير استخدامات بيانات الأرض وبيانات التواصل الاجتماعي في رصد العديد من المؤشرات الاجتماعية وغيرها من مؤشرات التي ترصد حركه وتوزيع وكثافه السكان.

اظهر التقرير العديد من الأدلة الدولية المستخدمة لوضع إطار محدد لاستخدام البيانات الضخمة في إنتاج المؤشرات الاجتماعية مثل "دليل استخدام بيانات الهاتف المحمول للإحصاءات الرسمية" والذي يوفر الإطار العالمي لاستخدام بيانات المحمول في إنتاج الإحصاءات الرسمية بدأيه من التأكد من جودة مصادر البيانات المستخدمة مرورا بخطوات إعدادها حتى إنتاج المؤشرات وعرض النتائج.

اظهر التقرير المجالات الإحصائية التي تم استخدام في دول جنوب اسيا والمحيط الهادي والتي تركزت في 4 مجالات تشمل: إحصاءات النقل والهجرة بأنواعها المختلفة، إحصاءات القوى العاملة، الإحصاءات الصحية بالإضافة الى رصد أهداف التنمية المستدامة، والتي تم فيها استخدام العديد من مصادر البيانات الضخمة السالفة الذكر والتي تتناسب مع طبيعة الإحصاءات وفقا للمفاهيم الدولية.

تطرق التقرير الى خبرات بعض الدول في تسهيل عمليه استخدام البيانات الضخمة ومنها شركات البيانات التي قام بها جهاز الإحصاء الاسترالي مع شركات الاتصالات للحصول على مزيد من بيانات التتبع المكاني للأفراد خلال جائحه كوفيد-19، كذلك الشركة بين جهاز الإحصاء الإندونيسي وشركه تيلكومسيل أكبر مشغل للهاتف المحمول في البلاد لإنتاج مؤشرات الهجرة والتنقل بين الأقاليم المختلفة، بالإضافة الى مذكرة التفاهم التي قام بها جهاز الإحصاء الكوري مع شركات المحمول الكورية من أجل إعداد أول تحليل لتحركات السكان على أساس بيانات الهاتف المحمول في عام 2017 .. وغيرها من الاتفاقات التي توفر الإطار التشريعي لاستخدام البيانات الضخمة في إنتاج الإحصاءات الرسمية بشكل او اخر.

وخلص التقرير الى اتجه معظم دول المنظمة لاستخدام البيانات الضخمة في انتاج إحصاءات النقل، الهجرة بشكل أكبر من الإحصاءات الحيوية وغيرها من الإحصاءات الاجتماعية وذلك لتوفر البيانات اللازمة بشكل يتيح إنتاج تلك الإحصاءات، واصى بضرورة تحديد المجالات ذات الأولوية لاستخدام البيانات الضخمة في الإحصاءات السكانية والاجتماعية وتجربة المشاريع الصغيرة التي إذا نجحت، كذلك ضرورة إعداد إطار قانونيا يتيح الاعتماد على البيانات الضخمة في إنتاج الإحصاءات الرسمية.

مدى الاستفادة من الدراسات السابقة

استفاد الباحث من الدراسات السابقة في التعرف على الجوانب الهامة في البحث العلمي كالإطار النظري للبحث والمنهج العلمي وأساليب التحليل الإحصائي المستخدم في البحث واستخلاص النتائج وتوظيف تلك الدراسات في تفسير نتائج البحث الميدانية، ورصد الخبرات الدولية في مختلف الأقاليم على مستوى العالم للاستفادة منها في رصد تحليل القدرات الوطنية لمواكبه الثورة العالمية في استخدام البيانات، كما أظهرت بعض العقبات الاجتماعية في استخدام البيانات الضخمة فيما يتعلق بالخصوصية وسريه البيانات الشخصية.

ثانياً: أوجه التشابه بين الدراسة الحالية والدراسات السابقة

اتفقت الدراسة الحالية مع معظم الدراسات السابقة في استخدام الاستبيان كأداة لجمع البيانات والاشتراك في منهج البحث الذي تم اختياره وهو المنهج الوصفي التحليلي، بالإضافة إلى أن معظم الدراسات السابقة أكدت على أهمية استخدام البيانات الضخمة في إنتاج الإحصاءات الرسمية بأنواعها وعلى راسها الإحصاءات السكانية والاجتماعية، كذلك رصد فرص وتحديات استخدام البيانات الضخمة بشكل عام وفى إنتاج الإحصاءات الرسمية بشكل خاص.

ثالثاً: أوجه الخلاف بين الدراسة الحالية والدراسات السابقة

تختلف الدراسة الحالية عن الدراسات السابقة في الهدف الرئيسي لها وهو محاولة تحديد فرص وتحديات استخدام البيانات الضخمة في إنتاج الإحصاءات الرسمية بجمهورية مصر العربية وما له من أثار في رفع جودة انتاج الإحصاءات الرسمية وتوفير مصادر بيانات غير تقليدية بتكلفه أقل كذلك توفير إحصاءات ومؤشرات اجتماعية بشكل أكثر سهوله وسرعه وأثار ذلك في عملية التخطيط والمتابعة والتقييم، كما تختلف هذه الدراسة عن الدراسات السابقة في العنوان والهدف والعينة ومكان التطبيق.

ومن هنا فإن هذه الدراسة تكمل الجوانب والمحاور التي لم تتطرق لها الدراسات السابقة مما يعزز الحاجة إليها، خاصة في حدودها الموضوعية التي تناولتها وما توصلت إليه من نتائج ومقترحات وتوصيات.

الإطار النظري للدراسة:

البيانات الضخمة:

تعد البيانات الضخمة أحد نتائج التقدم العلمي الهائل في شتى المجالات وتغلغل الرقمنه في شتى مجالات الحياة بما في ذلك الحياة الشخصية للمواطن بشكل عام دون النظر الى مستوى الدخل او حتى محل الإقامة، ومع تراكم هذا الكم الهائل من البيانات الناتج عن المعاملات اليومية في المجالات الاقتصادية والمالية حتى الاجتماعية أصبح لدى البشرية مخزون هائل من البيانات المتجددة والمتزايدة باستمرار وهو ما أصبح يطلق عليه البيانات الضخمة. بشكل عام لا يوجد اتفاق بين الأكاديميين على وضع تعريف محدد للبيانات الضخمة، وهنا يجدر الذكر على ان الحجم لسه هو السمة البارزة الوحيدة لتعريف هذا النوع من البيانات ولكن توجد سمات اخرى.

وتعددت التعريفات الخاصة بالبيانات الضخمة وتختلف فيما بينها في وصف الخصائص والسمات المميزة لها (Vs)،حيث عرفتها شركة جارتنر (Gartner Inc.2018) المتخصصة في أبحاث تقنية المعلومات بأنها "الأصول المعلوماتية كبيرة الحجم وسريعة التدفق وكثيرة التنوع ، والتي تتطلب طرق معالجة مجدية اقتصادياً ومبتكرة من أجل تطوير البصائر وطرق اتخاذ القرارات" ، كما عرفتها المنظمة الدولية (ISO,2017) بانها "مجموعة أو مجموعات من البيانات التي لها خصائص فريدة مثل الحجم ، السرعة ، التنوع ، التباين ، المصداقية وصحة البيانات" ، كما عرفها الاتحاد الدولي الاتصالات (ITU,2018)بأنها "مجموعة البيانات التي تتميز بالحجم الكبير والسرعة الفائقة والتنوع مقارنة بأنواع مجموعة البيانات المتعارف عليها."، ولكن يمكن القول بأن اكثر التعريفات شمولا هو التعريف الذى تقدمه مؤسسة TechAmerica Foundation,2012)بأنها "البيانات الضخمة هي مصطلح يصف كميات كبيرة من البيانات عالية السرعة والمعقدة والمتغيرة التي تتطلب أساليب وتقنيات متقدمة لتمكين استيعاب المعلومات وتخزينها وتوزيعها وإدارتها وتحليله "،ووفقا لهذا التعريف فإن البيانات الضخمة تتميز بما يسمى بال Vs الأربعة، وهي السرعة والحجم والمصداقية والتنوع.

السرعة: ويقصد بها سرعة إنتاج واستخراج البيانات لتغطية الطلب عليها، وهو الوقت الذي نستغرقه من لحظة وصول هذه البيانات إلى لحظة الخروج بالقرار بناء على استخدام هذه البيانات.

الحجم: يشير إلى تزايد حجم وسرعة نمو تلك البيانات وظهور مصادر جديدة لها، وهو ما يحدد قيمة وإمكانيات البيانات لكي تصنف من ضمن البيانات الضخمة، وبالتالي تحتاج إلى معالجات وأجهزة قادرة على التعامل معها.

المصداقية :يشير إلى موثوقية البيانات حيث أن المستفيدين يهتمون بالمعلومات حول جودة البيانات، وليس كل البيانات يمكن الاستفادة منها في خدمة المنظمة وصناعة القرار، لذا يتم التخلص من بعض البيانات غير المجدية.

التنوع :يشير ذلك إلى أن أكثر من 80 % من البيانات اليوم غير منظم أو غير مهيكله وعادة ما تكون تلك البيانات كثيرة جداً لإدارتها بشكل فعال، فالبيانات يتم استقبالها من خلال العديد من المصادر المختلفة للبيانات.

ويمكن التعبير عن تلك الخصائص والترابط فيما بينها كما موضح في الشكل التالي

شكل (1) خصائص البيانات الضخمة

المصدر: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS.

تصنيفات البيانات الضخمة:

يوجد العديد من التصنيفات للبيانات الضخمة نبرز هنا أبرز التصنيفات المتعارف عليها وهى كالتالي:

- التصنيف الأول: يقوم بتصنيف البيانات وفقا لطريقه تنظيمها حيث يمكن أن تقسم إلى ثالث أنواع هي:
  - بيانات مهيكله: وهي البيانات المنظمة في صورة جداول، أو قواعد بيانات تمهيدا لمعالجتها.
  - بيانات غير مهيكله: وهي تلك التي يولدها الأشخاص يوميا من كتابات، وصور، ونقرات على المواقع.
  - بيانات شبه مهيكله: وتعتبر نوعا من البيانات المهيكلة، الا أن البيانات لا تصمم في جداول أو قواعد بيانات[2]
- التصنيف الثاني يقوم بتصنيف أنواع البيانات الضخمة وفقا لمصادر البيانات[3]، على النحو التالي:
  - مصادر البيانات والمعلومات الناشئة عن إدارة برنامج ما، حكوميا أم لا، مثلً السجلات الطبية الإلكترونية.
  - مصادر تجارية أو معاملات ناشئة عن المعاملة بين كيانين، مثل معاملات بطاقات الائتمان عبر الإنترنت.
  - مصادر شبكة الاستشعار، مثل بيانات الكاميرا، والتصوير عن طريق الأقمار الصناعية، بيانات الاستشعار عن بعد
  - تتبع مصادر الأجهزة، مثل تتبع البيانات من الهواتف المحمولة والنظام العالمي لتحديد المواقع(GPS)
  - مصادر البيانات السلوكية، مثلا عمليات البحث عبر الإنترنت (حول منتج وأي نوع آخر من المعلومات).
  - مصادر بيانات الرأي، مثال التعليقات على وسائل التواصل الاجتماعي.
  - بيانات نظام المعلومات الجغرافية(GIS) ومعلومات من مصادر وأنواع مختلفة.

2- نبذة عن البيانات الضخمة http://bigdatainarabic.wordpress.com.

3-المجلس الاقتصادي والاجتماعي للأمم المتحدة، 2013

الإحصاءات الاجتماعية

هي استخدام الإحصاء لدراسة السلوك البشري والبيئات الاجتماعية. وتتكون بيانات الإحصاءات الاجتماعية من معلومات أو معرفة عن فرد أو كائن أو حدث.

تدعم الإحصاءات السكانية والاجتماعية قرارات السياسة القائمة على الأدلة والتي لها تأثير هائل على رفاهيتنا ونوعية حياتنا. تشمل الإحصاءات السكانية والاجتماعية كل جانب من جوانب حياتنا، من الصحة إلى التعليم، ومن التوظيف إلى الجريمة. على عكس الإحصاءات الاقتصادية والبيئية، لا يوجد إطار متفق عليه عالميًا للموضوعات أو الموضوعات التي تتضمن إحصاءات السكان والإحصاءات الاجتماعية. يشار إلى إحصاءات السكان على نطاق واسع باسم الإحصاءات الديموغرافية وتشمل الإحصاءات الحيوية مثل تلك التي تم تجميعها من أنظمة التسجيل المدني.

تم تحديد أربعة مواضيع للإحصاءات السكانية والاجتماعية على الصفحة الرئيسية لشعبة الإحصاءات في الأمم المتحدة – إحصاءات النوع الاجتماعي، وإحصاءات استخدام الوقت، إحصاءات الهجرة وإحصاءات الإعاقة.

اختبار الفروض

تناول الباحث في هذا الجزء من الدراسة العناصر الأساسية للدارسة الميدانية والذي يضم الفروض الخاصة بالدراسة وأهدافها والمستهدفين لإجراء الدارسة وكيفية جمع البيانات، كذلك متغيرات الدراسة والأساليب الإحصائية المستخدمة في التحليل واستخلاص النتائج واختبار الفروض

تصميم الدراسة الميدانية

تم إعداد استبيان لرصد واقع استخدام تحليل البيانات الضخمة في الإحصاءات الرسمية، يتكون من ثلاثة أقسام:

القسم الأول: خصائص المستجيب ويتكون من 5 أسئلة

القسم الثاني: استبيان حول موضوع الدراسة ويتكون من 20 سؤال

القسم الثالث: عقبات استخدام تحليل البيانات الضخمة ويتكون من 5 أسئلة

أستهدف الاستبيان الفئات الأساسية في المجتمع الإحصائي من منتجي البيانات الرسمية (الجهاز المركزي للتعبئة العامة والإحصاء) وبعض من مستخدمي البيانات من الأجهزة الحكومية (وزارة التخطيط...) والأجهزة الغير حكومية (منتدى البحوث الاقتصادية..) بالإضافة الى مجموعه من الأكاديميين بجامعة القاهرة.

مجتمع الدراسة

يتمثل مجتمع الدارسة في مجموعه من العاملين بالعمل الإحصائي بمختلف مستوياته انتاج/استخدام البيانات يمثلها الجدول التالي:

جدول (1) توزيع المجتمع على فئات الدراسة*

فئات الدراسة	العدد	%
جامع بيانات	85	56.7
مستخدم بيانات	40	26.7
أكاديمى	5	3.3
إدارى	20	13.3
الاجمالى	150	100.0

*المصدر : من إعداد الباحث

جمع البيانات

قام الباحث بجمع البيانات عن طريق إعداد نموذج الكترونى والتواصل مع الفئات المستهدفة بمختلف فئات مستويات العمل الإحصائي، ونظام المقابلة الشخصية مع قيادات العمل الإحصائي وبعض الأكاديميين.
بلغت عدد المستجيبين 160 استمارة لم يستكمل منهم الاستبيان عدد 10 استمارات حيث بلغت نسبة الاستجابة 93.8%وهي نسبة عالية يمكن الاعتماد عليها في التحليل واستخلاص المؤشرات الإحصائية اللازمة للتحقق من نتائج وفروض الدراسة محل البحث.

الخصائص الديموغرافية للمستجيبين:

يوضح الجدول التالي الخصائص الديموغرافية للمستجيبين وتبين ارتفاع نسبه المستجيبين الذكور المستجيبين عن الإناث حيث بلغت نسبه الذكور 56% من أجمالي المستجيبين، كذلك فان أكثر المستجيبين في الفئة العمرية من 30 حتى 39 حيث بلغت نسبتهم 42%، كما ان أغلب المستجيبين حاصلين على مؤهل البكالوريوس

جدول (3) الخصائص الديموغرافية للمستجيبين

المتغير	الفئات	التكرار	النسبة
النوع	ذكر	84	56.0
	أنثى	66	44.0
	الإجمالي	150	100.0
فئات السن	من 20 الى 29	35	23.3
	من 30 الى 39	54	36
	من 40 الى 49	44	29.3
	من 50 الى 59	17	11.3
	الإجمالي	150	100.0
المؤهل الدراسي	بكالوريوس	72	48.0
	دبلوم دراسات عاليا	34	22.7
	ماجستير	23	15.3
	دكتوراه	21	14.0
	الإجمالي	150	100.0

وتدل نتائج الجدول السابق على تنوع فئات المستجيبين لتمثيل جميع فئات المجتمع محل الدراسة، ولكن رغم هذا التنوع فإن جميع المستجيبين لديهم القدرة على فهم واستيعاب الاستبيان موضع الدراسة وخلفيتهم العلمية والعملية تؤهلهم لفهم أسئلة الاستبيان والإجابة عليها بدقة.

قياس الصدق والثبات:

يقصد بصدق أداة البحث : "دراسة مدل ملائمة الاستبيان المستخدم في قياس الغرض المصمم من أجله أو الظاهرة التي تسعى الباحث إلى معرفتها، ومدى قدرة هذه الأداة على قياس وتوفير المعلومات المطلوبة"، بينما يقصد بثبات

المقياس عادة أن يكون على درجة عالية من الدقة والإتقان والاتساق فيما يزودنا به من بيانات عن سلوك المبحوث، والاختبار التجريبي الثابت هو الذى يعطى نفِس النتائج (تقريباً) إذا طبق على نفس الأشخاص في فرصتين مختلفتين، كما يعرف ثبات المقياس بأنه المدى الذى يصل إليه المقياس في إعطاء قراءات متقاربة عند كل مرة يتم استخدامه فيها. والجدول التالي يوضح مقاييس الصدق والثبات كما يلي:

جدول (4) مقاييس الصدق والثبات لعبارات الاستبيان وفقا لاختبار الفا كرونباخ

السؤال	الثبات (%)	الصدق (قيمة الفا)	السؤال	الثبات (%)	الصدق (قيمة الفا)
هناك حاجه لاستخدام البيانات الضخمة لإنتاج الإحصاءات الرسمية	0.843	0.728	يتفق تحليل البيانات الضخمة مع طبيعة المجتمع المصري	0.777	0.728
يمكن تطبيق تحليل البيانات الضخمة في النظام الإحصائي الوطني	0.882	0.775	يعمل تحليل البيانات الضخمة على زيادة الأعباء الاقتصادية	0.830	0.728
القدرات الوطنية قادرة على استخدام البيانات الضخمة في انتاج الاحصاءات	0.723	0.785	يساهم تحليل البيانات الضخمة في التنبؤ بالمؤشرات الإحصائية	0.791	0.732
توجد العديد من العقبات تعوق تطبيق تحليل البيانات الضخمة	0.834	0.720	يؤدى تحليل البيانات الضخمة على سد فجوات البيانات	0.716	0.729
يمكن الاعتماد على البيانات الضخمة لإنتاج المؤشرات الإحصائية الوطنية	0.809	0.726	يساعد تحليل البيانات الضخمة في تلبيه رغبات مستخدمي البيانات	0.862	0.728
تحليل البيانات الضخمة يعمل على الغاء العمل الإحصائي التقليدية	0.761	0.722	يساعد تحليل البيانات الضخمة على زيادة جودة العمل الإحصائي	0.732	0.785
يمكن استخدام المؤشرات الناتجة عن البيانات الضخمة في التخطيط	0.832	0.719	تتفق تطبيق تحليل البيانات الضخمة مع مؤشرات التنمية المستدامة	0.891	0.759
يمكن استخدام تحليل البيانات الضخمة في إدارة الأزمات	0.862	0.734	يساعد تحليل البيانات الضخمة في المقارنات الدولية	0.873	0.753
تتفق البيانات الضخمة مع المبادئ الأساسية للإحصاءات الرسمية	0.842	0.779	المبادرات الدولية تساعد بصورة جيدة في تعظيم استخدام البيانات الضخمة في الأجهزة الإحصائية	0.916	0.771
تمتلك البيانات الضخمة القدرة على تحويل مفهوم البيانات وانتاج المؤشرات الوطنية	0.862	0.789	يمكن اعتماد تحليل البيانات الضخمة كأحد مصادر البيانات الأساسية للعمل الإحصائي في الأجل القريب	0.790	0.801
قيمة معامل الثبات		**0.834**	**قيمه معامل الصدق**		**0.751**

المصدر: مخرجات البرنامج الإحصائي

من الجدول السابق يتضح أن قيمة معاملي الثبات والصدق لأسئلة الاستبيان على التوالي (.834)، (.751) وهي نسبه عالية مما يؤكد إمكانية الاعتماد علي نتائج التحليلات والإحصائيات وتطبيقها على مجتمع الدراسة.

الإحصاءات الوصفية لمتغيرات الدراسة

سيتم في هذا الجزء التحليل الوصفي لمتغيرات الدراسة باستخراج مقاييس النزعة المركزية – ومقاييس التشتت – واختبار التوزيع الطبيعي لفروض الدراسة،

الجدول التالي يوضح مجموعة من الإحصاءات الوصفية لعبارات الاستبيان

جدول (5) الإحصاءات الوصفية لعبارات الاستبيان

رقم	العبارة	الوسط	الانحراف المعياري	معامل الاختلاف	الترتيب النسبي
1	هناك حاجه لاستخدام البيانات الضخمة لإنتاج الإحصاءات الرسمية	3.82	1.269	1.61	19
2	يمكن تطبيق تحليل البيانات الضخمة في النظام الإحصائي الوطني	4.19	0.787	0.619	9
3	القدرات الوطنية قادرة على استخدام البيانات الضخمة في انتاج الاحصاءات	4.05	0.577	0.333	4
4	توجد العديد من العقبات الفنية والتقنية في المجتمع الإحصائي الوطني تعوق تطبيق تحليل البيانات الضخمة	4.33	0.599	0.358	5
5	يمكن الاعتماد على البيانات الضخمة لإنتاج المؤشرات الإحصائية	3.56	0.698	0.488	7
6	تحليل البيانات الضخمة يعمل على الغاء العمل الإحصائي بالشكل التقليدية	4.23	0.572	0.327	1
7	يمكن استخدام المؤشرات الناتجة عن البيانات الضخمة في عمليات التخطيط	2.41	1.075	1.155	16
8	يمكن استخدام تحليل البيانات الضخمة في إدارة الأزمات	3.62	0.888	0.789	11
9	تتفق البيانات الضخمة مع المبادئ الأساسية للإحصاءات الرسمية	3.62	0.888	0.789	12
10	تمتلك البيانات الضخمة القدرة على تحويل مفهوم البيانات وانتاج المؤشرات الوطنية	4.23	0.572	0.327	2
11	يتفق تحليل البيانات الضخمة مع طبيعة المجتمع المصري	2.6	1.211	1.466	17
12	يعمل تحليل البيانات الضخمة على زيادة الأعباء الاقتصادية	3.33	1.029	1.06	14
13	يساهم تحليل البيانات الضخمة في التنبؤ بالمؤشرات الإحصائية	4.27	0.637	0.406	6
14	يؤدى تحليل البيانات الضخمة على سد فجوات البيانات الإحصائية	2.37	0.87	0.757	10
15	يساعد تحليل البيانات الضخمة في تلبيه رغبات مستخدمي البيانات	2.17	1.055	1.113	15
16	يساعد تحليل البيانات الضخمة على زيادة جودة العمل الإحصائي	2.6	1.211	1.466	18
رقم	العبارة	الوسط	الانحراف	معامل	الترتيب

		المعياري		الاختلاف	النسبي
17	تتفق تطبيق تحليل البيانات الضخمة مع مؤشرات التنمية المستدامة	4.23	0.572	0.327	3
18	المبادرات الدولية في تحليل البيانات الضخمة تساعد بصورة جيدة في تعظيم استخدام البيانات الضخمة في الأجهزة الإحصائية	3.82	1.269	1.61	20
19	يساعد تحليل البيانات الضخمة في المقارنات الدولية	3.62	0.888	0.789	13
20	يمكن اعتماد تحليل البيانات الضخمة كأحد مصادر البيانات الأساسية للعمل الإحصائي في الأجل القريب	3.56	0.698	0.488	8
	الإجمالي	3.72	0.857		

من الجدول السابق يتضح ما يلي

- احتل السؤال رقم (6) " تحليل البيانات الضخمة يعمل على الغاء العمل الإحصائي بالشكل التقليدية" على الترتيب النسبي الأول وفقا لمعامل الاختلاف (0.327) وبمتوسط حسابي (4.23)، وانحراف معياري بلغ (0.572).
- احتل السؤال رقم (18) " المبادرات الدولية في تحليل البيانات الضخمة تساعد بصورة جيدة في تعظيم استخدام البيانات الضخمة في الأجهزة الإحصائية" على الترتيب النسبي الأول وفقا لمعامل الاختلاف (1.61) وبمتوسط حسابي (3.82)، وانحراف معياري بلغ (1.269).
- أثبتت النتائج ارتفاع قيمه المتوسطات الى حد كبير حيث بلغ المتوسط الحسابي (3.72) وانحراف معياري قدرة (0.857)، وهو مستوى عال يدل على ميل المستجيبين الى الموافقة على عبارات الاستبيان بشكل عام.

اختبارات صحة الفروض

لاختبار صحة الفروض تم استخدام اختبار T-Testعن طريق الحزمة الإحصائية SPSS 24ويمكن عرض وتحليل نتائج التحليل الإحصائي بفرض العدم الخاص بالدراسة الميدانية خلال الجدول التالي:

الفرض الأول:

"لا توجد علاقة ذات دلالة احصائية بين أراء الفئات المستقصي منهم حول أهمية تحليل البيانات الضخمة"

يقوم الباحث باختبار وجود اختلافات معنويه ذات دلالة إحصائية بين أراء المستجيبين حول أهمية استخدام تحليل البيانات الضخمة في إنتاج الإحصاءات الرسمية، على ان يكون الفرض العدم هو عدم اختلافات معنويه بين أراء المشاركين في الاستبيان.

ويشمل هذا الفرض عدد 4 أسئلة، وكانت نتائج اختبار صحة الفرض في الجدول التالي:

جدول (6) نتائج اختبار t-testعند مستوى معنوية 0.05

لاختبار فرض الدراسة الاول

One-Sample Test						
95% Confidence Interval of the Difference		Mean Difference	Sig. (2-tailed)	Df	t	Q
Upper	Lower					
3.75	3.42	3.587	.000	149	42.341	1
3.56	3.17	3.367	.000	149	34.339	2
3.52	3.15	3.333	.000	149	35.840	3
3.58	3.22	3.400	.000	149	37.677	5
3.25	2.93	3.093	.000	149	38.445	9

قيمة t الجدولية عند مستوى دلالة < 0.5 ودرجات حرية 150 تساوى 1.96

من الجدول السابق يتضح ان قيمه ف المحسوبة تتراوح بين (34.339 و42.341) وهي أكبر من قيمة t الجدولية والتي تساوى (1.96) وبمستوى دلالة اقل من 0.05، مما يدل على ان غالبية اراء المستجيبة تدرك أهمية تحليل البيانات الضخمة والوعي بيها ادركا جيدا وهو ما يثبته التحليل الإحصائي السابق.

وبناء عليه يتم رفض الفرض العدم الأول وقبول الفرض البديل وينص على

"توجد علاقة ذات دلالة احصائية بين أراء الفئات المستقصي منهم حول أهمية تحليل البيانات الضخمة"

الفرض الثاني:

" لا توجد علاقة ذات دلاله إحصائية بين استخدام تحليل البيانات الضخمة وتطوير إنتاج المؤشرات الاجتماعية"

يقوم الباحث باختبار وجود اختلافات معنويه ذات دلالة إحصائية بين أراء المستجيبين حول أهمية استخدام تحليل البيانات الضخمة ورفع جودة العمل الإحصائي في إنتاج المؤشرات الاجتماعية، على ان يكون الفرض العدم هو عدم اختلافات معنويه بين أراء المشاركين في الاستبيان.

ويشمل هذا الفرض عدد 6 أسئلة، وكانت نتائج اختبار صحة الفرض في الجدول التالي:

جدول (7) نتائج اختبار t-test عند مستوى معنوية 0.05

لاختبار فرض الدراسة الثاني

One-Sample Test						
95% Confidence Interval of the Difference		Mean Difference	Sig. (2-tailed)	df	t	Q
Upper	Lower					
3.58	3.22	3.400	.000	149	37.677	5
3.81	3.47	3.640	.000	149	43.216	6
3.39	3.11	3.247	.000	149	45.485	10
3.38	3.02	3.200	.000	149	34.891	14
3.23	2.92	3.073	.000	149	39.370	15
3.41	3.13	3.267	.000	149	45.874	16

قيمة t الجدولية عند مستوى دلالة < 0.5 ودرجات حرية 150 تساوى 1.96

من الجدول السابق يتضح ان قيمه t المحسوبة تتراوح بين (34.891 و45.874) وهي أكبر من قيمة t الجدولية والتي تساوى (1.96) وبمستوى دلالة اقل من 0.05، مما يدل على ان غالبية اراء المستجيبين تدرك أهمية الدور التي سيقوم

به تحليل البيانات الضخمة في رفع كفاءة العمل الإحصائي وتعظيم الاستفادة من عملية التحول الرقمي في شتى مناحي الحياة.

وبناء عليه يتم رفض الفرض العدم الثاني وقبول الفرض البديل وينص على

" توجد علاقة ذات دلاله إحصائية بين استخدام تحليل البيانات الضخمة وتطوير إنتاج المؤشرات الاجتماعية"

الفرض الثالث:

" لا توجد علاقة ذات دلالة احصائية بين استخدام البيانات الضخمة وتقليل فجوات البيانات الاجتماعية"

يقوم الباحث باختبار وجود اختلافات معنويه ذات دلالة إحصائية بين أراء المستجيبين حول إمكانية استخدام تحليل البيانات الضخمة في تقليل فجوات البيانات الاجتماعية، على ان يكون الفرض العدم هو عدم اختلافات معنويه بين أراء المشاركين في الاستبيان.

ويشمل هذا الفرض عدد 3 أسئلة، وكانت نتائج اختبار صحة الفرض في الجدول التالي:

جدول (8) نتائج اختبار t-test عند مستوى معنوية 0.05

لاختبار فرض الدراسة الثالث

One-Sample Test						
95% Confidence Interval of the Difference		Mean Difference	Sig. (2-tailed)	df	T	Q
Upper	Lower					
3.57	3.23	3.400	.000	149	39.691	7
3.67	3.37	3.520	.000	149	47.365	8
3.14	2.78	2.960	.000	149	32.644	13

قيمة t الجدولية عند مستوى دلالة < 0.5 ودرجات حرية 150 تساوى 1.96

من الجدول السابق يتضح ان قيمه t المحسوبة تتراوح بين (32.644 و47.365) وهي أكبر من قيمة t الجدولية والتي تساوى (1.96) وبمستوى دلالة اقل من 0.05، مما يدل على ان غالبية اراء المستجيبين توكد على إمكانية استخدام نتائج تحليل البيانات الضخمة في عمليات التخطيط بشكل عام والتخطيط لا جل التنمية بشكل خاص، كما يمكن الاستفادة منها في توقع معدلات التغير وإدارة الأزمات.

وبناء عليه يتم رفض الفرض العدم الثالث وقبول الفرض البديل وينص على

" توجد علاقة ذات دلالة احصائية بين استخدام البيانات الضخمة وتقليل فجوات البيانات الاجتماعية"

الفرض الرابع:

" لا توجد علاقة ذات دلالة احصائية بين استخدام تحليل البيانات الضخمة واعدد خطط التنمية الاجتماعية"

يقوم الباحث باختبار وجود اختلافات معنويه ذات دلالة إحصائية بين أراء المستجيبين حول أهمية استخدام تحليل البيانات الضخمة في المساعدة في إعداد وتحقيق خطط التنمية الاجتماعية، على ان يكون الفرض العدم هو عدم اختلافات معنويه بين أراء المشاركين في الاستبيان.

ويشمل هذا الفرض عدد 4 أسئلة، وكانت نتائج اختبار صحة الفرض في الجدول التالي:

جدول (9) نتائج اختبار t-test عند مستوى معنوية 0.05 لاختبار فرض الدراسة الرابع

One-Sample Test						
95% Confidence Interval of the Difference		Mean Difference	Sig. (2-tailed)	df	t	Q
Upper	Lower					
3.92	3.62	3.767	.000	149	50.001	17
3.78	3.39	3.587	.000	149	36.453	18
3.66	3.36	3.507	.000	149	45.717	19
3.61	3.19	3.400	.000	149	31.284	20

قيمة t الجدولية عند مستوى دلالة < 0.5 ودرجات حرية 150 تساوى 1.96

من الجدول السابق يتضح ان قيمه ف المحسوبة تتراوح بين (31.284 و50.001) وهي أكبر من قيمة t الجدولية والتي تساوى (1.96) وبمستوى دلالة اقل من 0.05، مما يدل على ان غالبية اراء المستجيبين تدرك أهمية الدور التي سيقوم به تحليل البيانات الضخمة في المساعدة على تحقيق أهداف التنمية المستدامة.

وبناء عليه يتم رفض الفرض العدم الرابع وقبول الفرض البديل وينص على

" لا توجد علاقة ذات دلالة احصائية بين استخدام تحليل البيانات الضخمة واعدد خطط التنمية الاجتماعية"

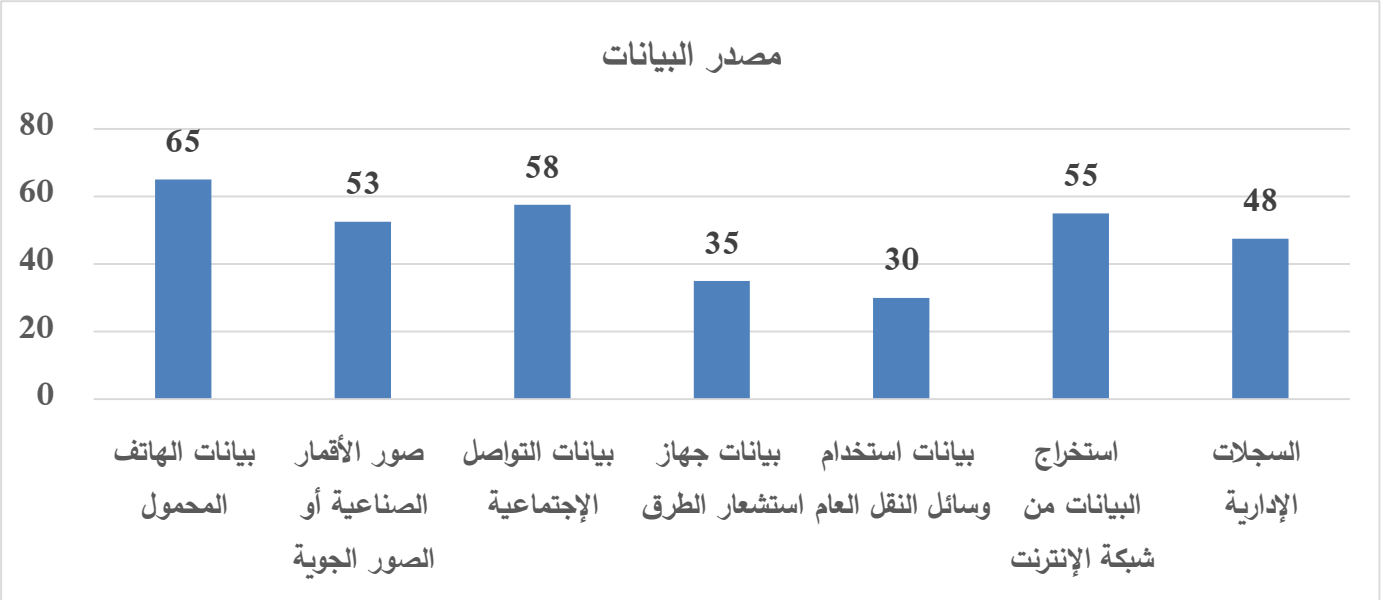
تحليل أراء المشاركين حول أهم العقبات التي تواجه استخدام تحليل البيانات الضخمة وطنيا:

في هذا الجزء يتناول الباحث تحليل للجزء الثالث من الاستبيان حول أراء المشاركين حول عقبات استخدام البيانات الضخمة وطنيا في مجال الإحصاءات الرسمية، وذلك عن طريق مجموعه مع الرسم البيانية والجداول التكرارية التي تبين تفاوت أراء المشاركين حول المواضيع محل الدارسة.

أولا: ما هي المصادر الرئيسية للبيانات الضخمة التي يمكن استخدمها

يوضح شكل (2) أراء المشاركين في الاستبيان حول مصادر البيانات الضخمة التي يمكن الاستفادة منها في تحليل البيانات الضخمة. فنجد ان المشاركين بنسبة (65%) يرون ان بيانات المحمول أحد اهم مصادر البيانات الضخمة التي يمكن الاستفادة منها إحصائيا يليها بيانات التواصل الاجتماعي واستخراج البيانات من الانترنت بنسب (58، 53) % على التوالي.

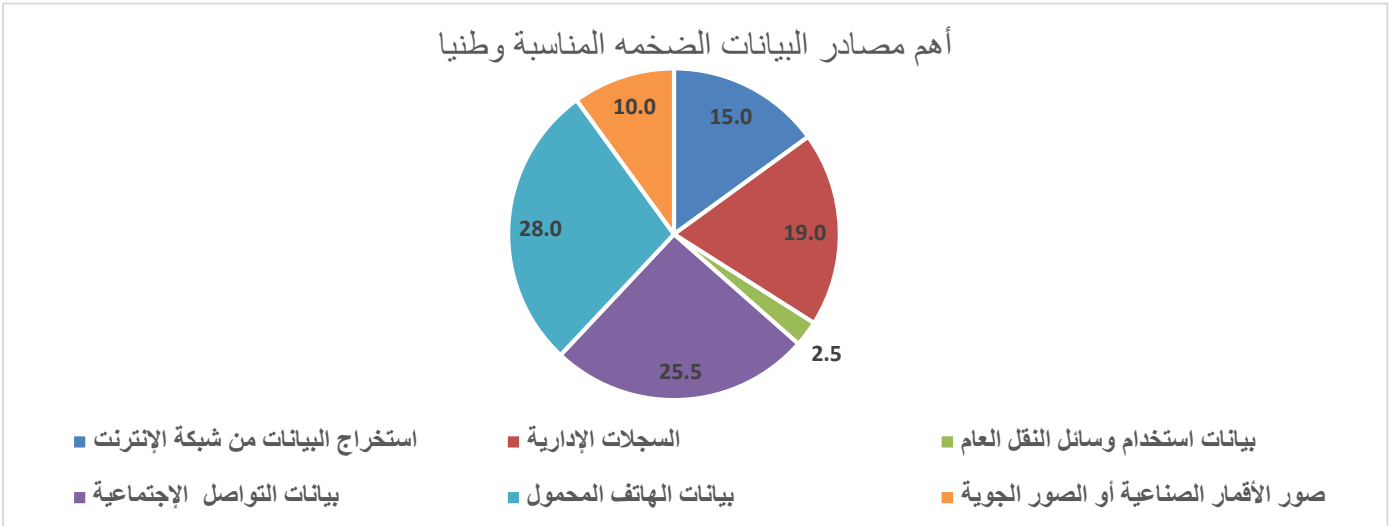
شكل (2) مصادر البيانات الضخمة



أهم مصادر البيانات الضخمة الملائمة للمجتمع المصري

يوضح شكل (3) اراء المستجيبين حول أهم مصادر البيانات الضخمة التي تتناسب بشكل كبير مع خصائص المجتمع المصري، ويمكن تقبل نتائجه بشكل كبير لفئه كبيرة من مستخدمي البيانات، فنجد ان ما يقرب من (28%)من المستجيبين يرى بيانات الهاتف المحمول هي ألاهم في المجتمع المصري يليها بيانات التواصل الاجتماعي يليها السجلات الإدارية بنسب(25.5 ، 19 %) على التوالي.

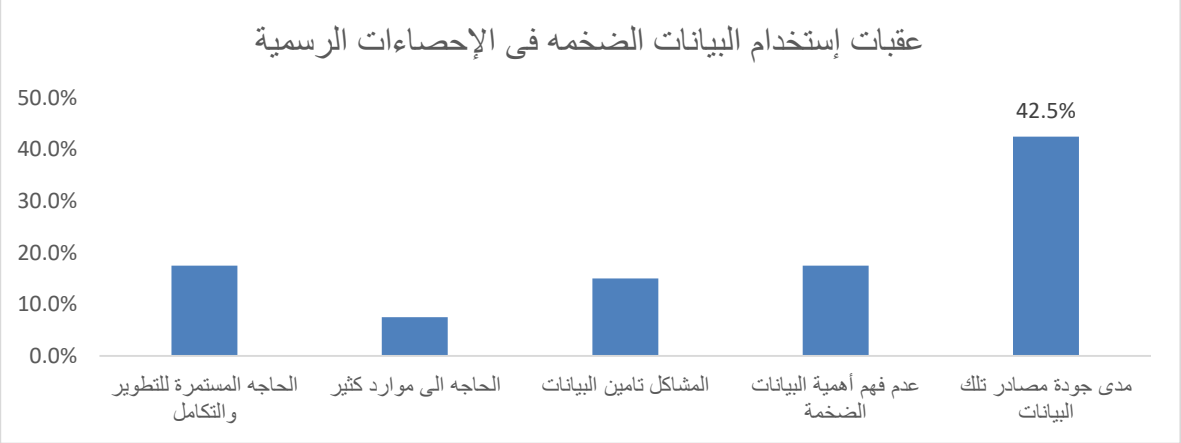
شكل (3) أكثر مصادر البيانات الضخمة ملائمه للمجتمع المصري



أهم التحديات والعقبات التي تعيق استخدام مصادر البيانات الضخمة في إنتاج الإحصاءات الوطنية

يوضح شكل (4) أن (42.5%)من المشاركة يجد ان مشكله المصداقية او التأكيد من جودة مصادر البيانات الضخمة هو العقبة الأساسية في استخدام تحليل البيانات الضخمة في إنتاج الإحصاءات الرسمية

شكل (4) أهم عقبات استخدام البيانات الضخمة في الإحصاءات الرسمية

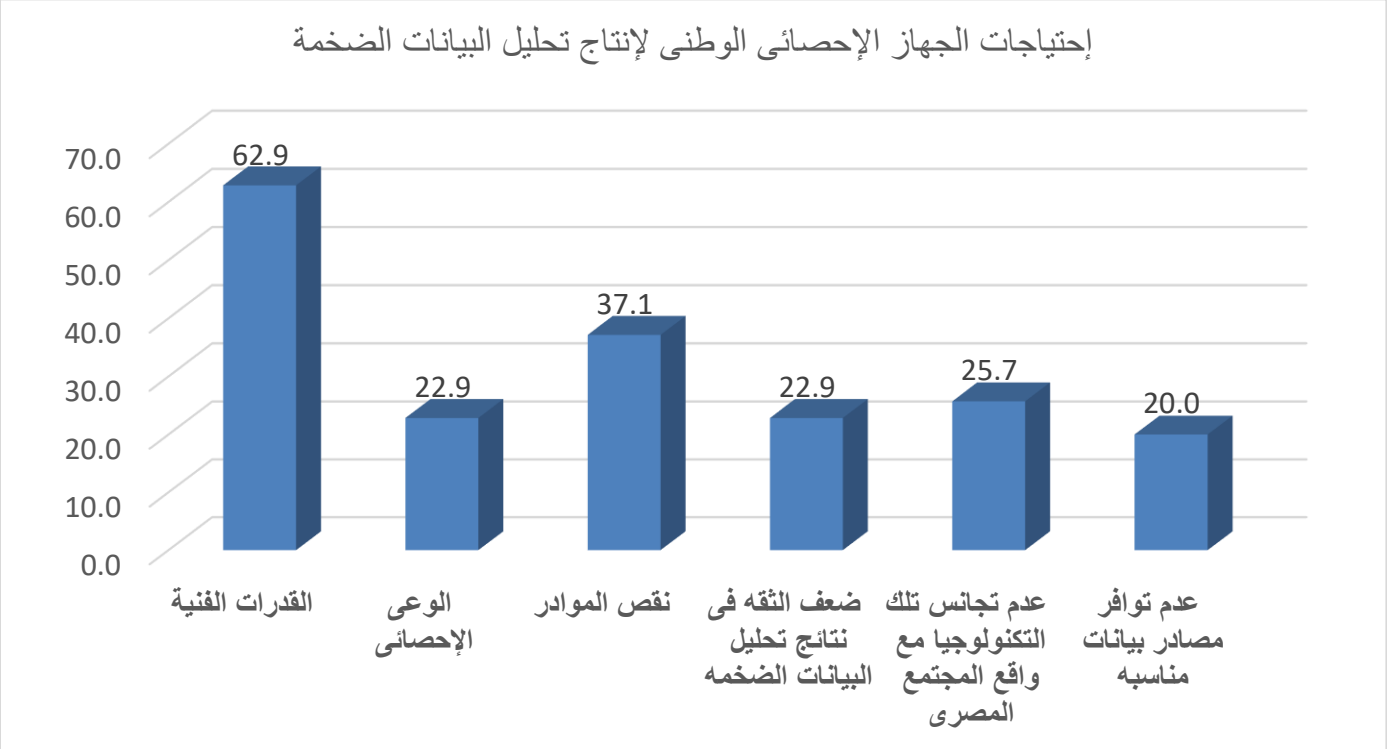


برأيك ما الذي ينقص جهاز الإحصاء الوطني لاستخدام البيانات الضخمة لإنتاج المؤشرات الوطنية

كما يظهر في الشكل التالي أن معظم المشاركين في الاستبيان بنسبة (62.9%) يرى أن القدرات الفنية أحد أهم التحديات التي توجها الأجهزة الإحصائية بشكل عام والجهاز الوطني بشكل خاص في استخدام تحليل البيانات الضخمة لإنتاج المؤشرات الإحصائية بشكل عام. يليها نقص الموارد بنسبة تصل الى 37.1%.

شكل (5)

أهم احتياجات الجهاز الإحصائي الوطني لاستخدام تحليل البيانات الضخمة في إنتاج المؤشرات الإحصائية



التحليل الرابعى (SWAT) للنظام الإحصائي في استخدام البيانات الضخمة لإنتاج المؤشرات الوطنية

في هذا الجزء يرصد الباحث تحليل مفصل للمجتمع الإحصائي وذلك من خلال تحليل عوامل البيئة الداخلية (نقاط القوة والضعف) والبيئة الخارجية (الفرص والتحديات) وذلك من خلال أراء المشاركين في البحث وخبرات الباحث.

البيئة الداخلية

نقاط القوة

- اتباع أحدث الأساليب التكنولوجيا في العمل الإحصائي بمراحله
- البنية التكنولوجية الإحصائية الجيدة
- المصدر الرسمي للبيانات
- تطبيق المفاهيم العالمية في العمل الإحصائي
- اتفاقيات التعاون مع شركاء العمل الإحصائي
- إتاحة البيانات

نقاط الضعف

- نقص الموارد
- نقص الخبرات
- التركيبة النوعية والعمرية للعاملين
- المستوى التعليمي
- مصدر البيانات الملائم

البيئة الخارجية

الفرص

- استراتيجية العمل الإحصائي
- التحول الرقمي
- فجوات البيانات
- الشركات العالمية
- السجلات الادارية

التحديات

- الوعي الإحصائي
- امن المعلومات
- مصداقية البيانات
- الحرك المستمر
- الازمات العالمية
- المفاهيم الإحصائية

نتائج البحث

تمثل البيانات الضخمة فرصة لتوليد منتجات معلوماتية في الوقت الفعلي، بينما تجلب الإحصاءات الرسمية عمق التفاصيل والتمثيل من خلال الممسوحات والتعدادات التي تم التحقق من صحتها. يمكن أن يؤدي مزج الاثنين إلى نتائج مثيرة، ولكن جني الفوائد ليس بالأمر السهل.

تعد البيانات الضخمة أحد نتائج التطور الحضاري للسلوك البشرى فأصبح كل شيء يتعامل من خلال البيانات وعن طريق التطبيقات المنتشرة في كل جزء في الكون، وفى ظل الظروف الاقتصادية والاجتماعية التي تطلب المزيد من التدبير في استهلاك الموارد المتاحة، ناهيك عن القيود التي تفرض على عمليات جمع البيانات من إجراءات صحية وقائية، لذا فمن الضروري الاستفادة من كل الموارد المتاحة وتعظيم الاستفادة من كل ما نملكه من كنوز في شتى المجالات.

ما سبق تبين بما لا يترك مكان للشك أهيمه تحليل البيانات الضخمة في إنتاج المؤشرات الإحصائية عند جميع الفئات العاملة بالمجال الإحصائي باختلاف قناعتهم وادوارهم في دائرة العمل الإحصائي، وذلك لما لها من دور كبير في سد فجوات البيانات والمؤشرات الوطنية بشكل عام والاجتماعية بشكل خاص.

توفر أساليب البيانات الضخمة صورة متكاملة للخصائص الاجتماعية للسكان بدمج العديد من المؤشرات المختلفة في إطار واحد يعمل على المساهمة في تحديد احتياجات السكان بشكل أكثر شمولا ويوفر كم هائل من المؤشرات الاجتماعية المتسقة والمترابطة فيما بينها.

تملك أساليب تحليل البيانات الضخمة وفقا لما ورد من أراء المشاركين القدرة على تطوير العمل الإحصائي رفع جودة المؤشرات الاجتماعية اما بإنتاج مؤشرات مفقودة أو تطوير أساليب حساب المؤشرات الحالية، والذي بدوره يساعد بشكل كبير في تطوير عمليه جمع البيانات الى شكل أكثر مرونة واقل تكلفه والذي بدور يحدث ثورة في دائرة العمل الإحصائي ويعظم الاستفادة من البيانات في إعداد الخطط الاجتماعية.

يمكن للبيانات الضخمة المساعدة في رصد معدلات التنمية الاجتماعية بشكل أكثر واقعيه وبصورة دورية دقيقه وفقا لأحدث البيانات وهو ما يوفر أفضل الأساليب للمتابعة والتقييم التي تمثل العنصر الرئيسي لضمان نجاح برامج التنمية وصلها لمستحقيه، بشكل سريع ودقيق دون إضافة أي أعباء إضافية على النظام الإحصائي.

تمتلك البيانات الضخمة ميزة تنافسيه لا تتوفر في العديد المؤشرات الإحصائية، حيث يتح استخدام تحليل البيانات الضخمة المزيد من فرص المقارنات الدولية انطلاقا من المفهوم العالمي للبيانات والمصادر المفتوحة للبيانات التي تتشبه بشكل كبير في العديد من البلدان.

تملك أساليب تحليل البيانات الضخمة وفقا لما ورد من أراء المشاركين القدرة على المساعدة في دراسات التنبؤ والتوقع بقيم المؤشرات الإحصائية المختلفة على كافة المستويات الجغرافية، والذي بدوره يساهم بشكل كبير في صياغة الخطط المناسبة لمواكبه التطورات المستقبلية المتلاحقة.

ولكن نتائج تحليل البيانات الضخمة يجد صعوبة في تقبل نتائجه في ظل ظروف اجتماعية مقعدة بعض الشي في ظل انخفاض معدلات الوعى الإحصائي، كما تعانى أساليب تحليل الضخمة من حدثتها بعض الشيء في مجتمعنا

والذى يتطلب المزيد من المجهودات حول تعلم تلك الأساليب الجديدة، نهيك عن الموارد التي يجب توفيرها بشكل أولى وتطويرها بشكل مستمر، ولكن بشكل أولى يمكن الاستعانة ببعض المبادرات الدولية والاهتمام بالدراسات الإقليمية حول استخدام البيانات الضخمة.

توصيات البحث

في ضوء النتائج السابقة يوصى الباحث:

- العمل على الاستفادة من المبادرات العالمية في استخدام البيانات الضخمة في رفع القدرات الفنية لتعظيم الاستفادة من تحليل البيانات الضخمة.
- ضرورة الإسراع في تطبيق استراتيجية العمل الإحصائي الوطني لإضافة الشكل القانوني على عمليه تكامل البيانات، مع التأكيد على إشراك جميع شركاء العمل الإحصائي.
- العمل على إعداد الدراسات المستفيضة في انتاج الإحصاءات الاجتماعية باستخدام أساليب تحليل البيانات الضخمة بشكل تجريبي على بعض المؤشرات وتعميمها بعد التأكد من جودتها.
- الاستفادة من تحليل البيانات الضخمة في إعداد خطط التنمية الاجتماعية.

التوجهات المستقبلية للباحث

- دور البيانات الضخمة في دعم أهداف التنمية المستدامة.
- إعداد نموذج لإنتاج بعض المؤشرات الاجتماعية باستخدام البيانات الضخمة

المراجع

أولا: المراجع العربية

- إعلان الدوحة بشأن ثورة البيانات في المنطقة العربية

https://www.arabdevelopmentportal.com/ar/news/doha-declaration-data-revolution-arab-region

- الأمم المتحدة. أهداف التنمية المستدامة.

www://https.un.org/sustainabledevelopment/ar/sustainable-development-goals/

- الامم المتحدة. المجلس الاقتصادي والاجتماعي البيانات الضخمة وتحديث النظم الإحصائية: تقرير الامين العام

/https://unstats.un.org/unsd/statcom/doc14-11.

- تقرير الفريق العالمي المعني باستخدام البيانات الضخمة لأغراض الإحصاءات الرسمية

https://unstats.un.org/unsd/statcom/48th.../2017-7-BigData-A.pdf.

- حارب، ياسر. البيانات الضخمة .ramadanahla.com
- نبذة عن البيانات الضخمة، https://bigdatainarabic.wordpress.com/
- استخدام البيانات الضخمة لقياس مؤشرات أهداف التنمية المستدامة في المنطقة العربية: التحديات والفرص-الاسكوا- مايو 2021.
- تقرير ركتن للإعلام، (2018)، "البيانات الضخمة ما أهميتها وما أهمية الاستثمار في تحليلها، وكيف ستؤثر في حياتنا وقرارتنا؟
- (ملياني فتيحة، 2019) بعنوان "البيانات الضخمة: الفرص، التحديات، ومجالات التطبيق"
- (مقناني 2019) بعنوان "دور البيانات الضخمة في دعم التنمية المستدامة بالدول العربية"
- جمال بن مطر السالمي، (2018)، "البيانات الضخمة ودورها في دعم اتخاذ القرار والتخطيط الاستراتيجي.
- سامر مظهر قنطقجي، (2014)، سوق البيانات الضخمة ومفاهيم جديدة"، مجلة الاقتصاد الإسلامي العالمية ، سوريا ، ديسمبر.
- العصيمي، عايد عبد الله. (2015) المسؤولية الاجتماعية للشركات نحو التنمية المستدامة. عمان: اليازوري.
- كيف تستطيع البيانات الضخمة دعم التنمية في إفريقيا.

ثانيا: المراجع الأجنبية:

- (ECLAC,2020) "An assessment of big data for official statistics in the Caribbean
- (ADB ,2020) "MAPPING POVERTY THROUGH DATA INTEGRATION AND ARTIFICIAL INTELLIGENCE"
- Ann C. Dzuranin, and Irina Malaescu, (2016), "The Current State and Future direction of IT Audit: Challenges and Opportunities", Journal of Information Systems, Vol. (30), No. (1), Spring.

- Deniz Appelbaum, (2016), "Securing Big Data Provenance for Auditors: The Big Data Provenance Black Box as Reliable Evidence", Journal of Emerging Technologies in Accounting, Vol. (13), No. (1), Spring.
- S. Alkire, et al. 2019. Changes over time in the global Multidimensional Poverty Index: A ten-country study, OPHI MPI Methodological Note 48. Oxford Poverty and Human Development Initiative, University of Oxford.
- - Gartner Inc, (2018), "Big Data", Available at: https://www.gartner.com/glossary/big-data,accessed 18 March.
- Hilbert Martin, (2015), "Big Data for Development: A Review of Promises and Challenges. Development Policy Review". Martinhilbert.net. Retrieved 7 October.
- Kantmar, (2018), "Analyzing big data: the path to competitive advantage", Electronic Copy Available at: http://hosteddocs. ittoolbox.com, accessed 28 October 2017
- Kitchin R., (2014), "The Data Revolution: Big Data, Open Data", Data Infrastructures & Their Consequences. London: SAGE Publications.
- Big Data for the SDGs Country examples in compiling SDG indicators using non-traditional data sources, ESCAP Statistics Division
- Reflections on Big Data & the Sustainable Development Goals: Measuring & Achieving Development Progress in the Big Data Era.
- Big data's great promise for development will need both human and technical capacity building, reports *Jan Piotrowski*.
- J. E. Steele, P. R. Sundsøy, C. Pezzulo, V. A. Alegana, T. J. Bird, J. Blumenstock, J. Bjelland, K. Engø-Monsen, Y.-A. de Montjoye, A. M. Iqbal, K. N. Hadiuzzaman, X. Lu, E. Wetter, A. J. Tatem and L. Bengtsson, "Mapping poverty using mobile phone and satellite data" (Journal of the Royal Society Interface, 2017).
- M. S. Andreano, R. Benedetti, F. Piersimoni, P. Postiglione and G. Savio, "Sampling and modelling issues using big data in now-casting", in New Statistical Developments in Data Science (eds. Verde R., Ferrari F., Petrucci A. and Racioppi F.), Springer Verlag, 2019.
- C. D. Elvidge, F.-C. Hsu, T. Ghosh, and M. Zhizhin, "World tour of COVID-19 impacts on nighttime lights", Earth Observation Group, Payne Institute for Public Policy Colorado School of Mines, 21 April 2020. https://payneinstitute.mines.edu/wp-content/uploads/sites/149/2020/04/World-Tour-of-COVID-19-Impacts-onNighttime-Lights-3.pdf.
- Jayaraman, R. (2018). Big Data in Decision Making Processes. Retrieved from https://m.esmt.org/ system/files_force/leporello-big-web_0.pdf? download=1&download=1.
- M. S. Andreano, R. Benedetti, F. Piersimoni and G. Savio, "Mapping poverty indices for Latin American and the Caribbean countries through satellite remote sensing" (Social Indicators Research, Springer, 2020). S. Cecchini, G. Savio and V. Tromben, "Shedding light on territorial poverty in Chile", submitted to Regional Science Policy and Practice, Wiley, 2020.
- J.R. Albert, et al. 2019. Readiness of National Statistical Systems in Asia and the Pacific for Leveraging Big Data to Monitor the SDGs. ADB Brief No. 106. Manila
- https://www.socialsciences.manchester.ac.uk/social-statistics/study/what-is-social-statistics/

الملاحق

I – الإستبيان

إستطلاع رأى حول

البيانات الضخمة ودورها في تحسين جودة المؤشرات الاجتماعية

إعداد
أحمد جمال محمد رمضان

إشراف
د.أن نصير

القسم الأول : خصائص المستجيب

م	السؤال	الاجابة
1	الأسم (إختيارى)	
2	السن	1- من 20 الى 29 2- من 30 الى 39 3- من 40 الى 49 4- من 50 الى 59
3	النوع	1- ذكر 2- أنثى
4	المؤهل الدراسى	1- بكالريوس 2- دبلوم دراسات عاليا 3- ماجستير 4- دكتوراه
5	طبيعة العمل	1- جامع بيانات 2- مستخدم بيانات 3- أكاديمى 4- إدارى

القسم الثانى : الإستبيان

م	النص	اتفق تمام	اتفق	محايد	غير موافق	غيرموافق تمام
1	هناك حاجه لإستخدام البيانات الصخمه لإنتاج الإحصاءات الرسمية					
2	يمكن تطبيق تحليل البيانات الضخمه في النظام الإحصائى الوطنى					
3	القدرات الوطنيه قادرة على تطبيق تحليل البيانات الضخمه فى انتاج الاحصاءات الرسمية					
4	توجد العديد من العقبات الفنية والتقنيه في المجتمع الإحصائى الوطنى تعوق تطبيق تحليل البيانات الضخمة					
5	يمكن الاعتماد على البيانات الضخمه لإنتاج المؤشرات الإحصائيه الوطنية					
6	تحليل البيانات الضخمه يعمل على الغاء العمل الإحصائى بالشكل التقليدية					

م	النص	اتفق تمام	اتفق	محايد	غير موافق	غيرموافق تمام
7	يمكن استخدام المؤشرات الناتجه عن البيانات الضخمه فى عمليات التخطيط					
8	يمكن استخدام تحليل البيانات الضخمه في إحصائيات إدارة الأزمات					
9	تتفق البيانات الضخمه مع المبادى الأساسية للإحصاءات الرسمية					
10	تمتلك البيانات الضخمه القدرة على تحويل مفهوم البيانات وانتاج المؤشرات الوطنية					
11	يمكن استخدام البيانات الضخمه لإنتاج الإحصاءات الرسمية					
12	يعمل تحليل البيانات الضخمه على زيادة الأعباء الاقتصادية على المجتمع					
13	يساهم تحليل البيانات الضخمه في التنبؤ بالمؤشرات الإحصائية					
14	يؤدى تحليل البيانات الضخمه على سد فجوات البيانات الإحصائية					
15	يساعد تحليل البيانات الضخمه في تلبيه رغبات مستخدمى البيانات					
16	يساعد تحليل البيانات الضخمه على زيادة جودة العمل الإحصائى					
17	تتفق تطبيق تحليل البيانات الضخمه مع مؤشرات التننمية المستدامة					
18	المبادرات الدولية في تحليل البيانات الضخمه تساعد بصورة جيدة في تعظيم إستخدام البيانات الضخمه في الأجهزة الإحصائية					
19	يساعد تحليل البيانات الضخمه في المقارنات الدولية					
20	يمكن إعتماد تحليل البيانات الضخمه كاحد مصادر البيانات الأساسية للعمل الإحصائى في الأجل القريب					

القسم الثالث : عقبات إستخدام تحليل البيانات الضخمه

م	السؤال	الاجابة
301	برايك ما هي المصادر الرئيسية للبيانات الضخمة التى يمكن استخدمها	(1) بيانات الهاتف المحمول (2) صور الأقمار الصناعية أو الصور الجوية (3) بيانات التواصل الإجتماعية (4) بيانات جهاز استشعار الطرق (5) بيانات استخدام وسائل النقل العام (6) استخراج البيانات من شبكة الإنترنت
302	برايك ما هي أكثر تلك المصادر ملائمة مع خصائص المجتمع المصرى	(1) بيانات الهاتف المحمول (2) صور الأقمار الصناعية أو الصور الجوية (3) بيانات التواصل الإجتماعية (4) بيانات جهاز استشعار الطرق (5) بيانات استخدام وسائل النقل العام (6) استخراج البيانات من شبكة الإنترنت
303	ما هي أهم التحديات والعقبات التي تحول دون استخدام مصادر البيانات الضخمة في إنتاج الإحصاءات الوطنية	(1) عدم فهم أهمية البيانات الضخمة (2) مدى جودة مصادر تلك البيانات (3) الحاجه الى موارد كثير (4) الحاجه المستمرة للتطوير والتكامل (5) المشاكل الأمنية
304	برايك ما الذى ينقص جهاز الإحصاء الوطنى لإستخدام البيانات الضخمه لإنتاج المؤشرات الوطنيه	(1) القدرات الفنية (2) نقص الموادر (3) عدم توافر مصادر بيانات مناسبه (4) ضعف الثقه فى نتائج تحليل البيانات الضخمه (5) الوعى الإحصائى (6) عدم تجانس تلك التكنولوجيا مع واقع المجتمع المصرى (7) غير ذلك