baseera بصيرة
EUROPEAN UNION
الاتحاد الأوروبي
المجلس القومي للمرأة

# استخدام البيانات الضخمة في توفير مؤشرات حول تمكين المرأة

د. حنان جرجس
نائب الرئيس التنفيذى لمركز بصيرة
د. عمار أحمد
أستاذ بكلية الدراسات العليا للبحوث الإحصائية بجامعة القاهرة

24
2030 استراتيجية المرأة WOMEN'S STRATEGY

baseera بصيرة
EUROPEAN UNION
الاتحاد الأوروبي
المجلس القومي للمرأة

# استخدام البيانات الضخمة في توفير مؤشرات حول تمكين المرأة

د. حنان جرجس
نائب الرئيس التنفيذى لمركز بصيرة
د. عمار أحمد
أستاذ بكلية الدراسات العليا للبحوث الإحصائية بجامعة القاهرة

24
2030 استراتيجية المرأة WOMEN'S STRATEGY

"الآراء الواردة في هذا التقرير لا تعكس بالضرورة وجهة نظر المجلس القومي للمرأة"

## جدول المحتويات


١- مقدمة ١

٢- ما هي البيانات الضخمة؟ ٢

٣- تجميع وتحليل البيانات الضخمة ٤

٤- البيانات الضخمة من أجل التنمية ٧

٥- البيانات الضخمة وتمكين المرأة ٨

٦- تحديات استخدام البيانات الضخمة في حساب مؤشرات تمكين المرأة ١١

٧- التوصيات ١٣

٨- المصادر ١٥

## ١- مقدمة

أطلقت مصر الاستراتيجية الوطنية لتمكين المرأة المصرية في عام ٢٠١٦، وتعتمد الاستراتيجية على ٤ محاور رئيسية هي التمكين الاقتصادي، والتمكين الاجتماعي، والتمكين السياسي، والحماية. وقد حددت وثيقة الاستراتيجية عدداً من المؤشرات لاستخدامها في متابعة تنفيذ الاستراتيجية. وفي عام ٢٠١٧، تم إطلاق مرصد المرأة المصرية ليعمل كأداة لمتابعة تنفيذ الاستراتيجية، وإلى جانب مؤشرات المتابعة التي حددتها الاستراتيجية، يعمل المرصد على توفير مجموعة من المؤشرات المساندة وهي مؤشرات ليست ضمن مؤشرات متابعة الاستراتيجية لكنها مؤشرات تقيس بعض العوامل التي قد يكون لها تأثير مباشر أو غير مباشر على تمكين المرأة.

ويعد نقص البيانات المطلوبة لمتابعة تمكين المرأة في مصر من أهم التحديات التي تواجه تمكين المرأة في مصر، ويتطلب توفير البيانات توافر موارد مالية وبشرية ضخمة.

وتشكل البيانات الضخمة فرصة كبيرة لتوفير البيانات من مصادر غير تقليدية وبتكلفة قد تكون أقل.

ويمكن توظيف البيانات الضخمة المتاحة من المصادر المختلفة في عمليات المتابعة والتقييم الخاصة بتنفيذ الاستراتيجية الوطنية لتمكين المرأة.

إلا أن هناك العديد من المحددات التي يجب وضعها في الاعتبار عند استخدام البيانات الضخمة في توفير مؤشرات حول تمكين المرأة، كما لا يخلو استخدام تلك البيانات من تحديات تحتاج إلى مزيد من البحث والجهود للتغلب عليها.

وتتعرض هذه الورقة في أقسامها التالية لفكرة استخدام البيانات الضخمة في توفير مؤشرات حول تمكين المرأة بمزيد من التفصيل، حيث يعرض القسم الثاني تعريف البيانات الضخمة، ويتطرق القسم الثالث لتجميع وتحليل البيانات الضخمة، بينما القسم الرابع يعرض فكرة استخدام البيانات الضخمة من أجل التنمية، ويركز القسم الخامس على استخدام البيانات الضخمة في تمكين المرأة حيث يعرض أمثلة لذلك، ويستعرض القسم السادس تحديات استخدام البيانات الضخمة في حساب مؤشرات تمكين المرأة، وتنتهي الورقة بمجموعة من التوصيات في القسم السابع.

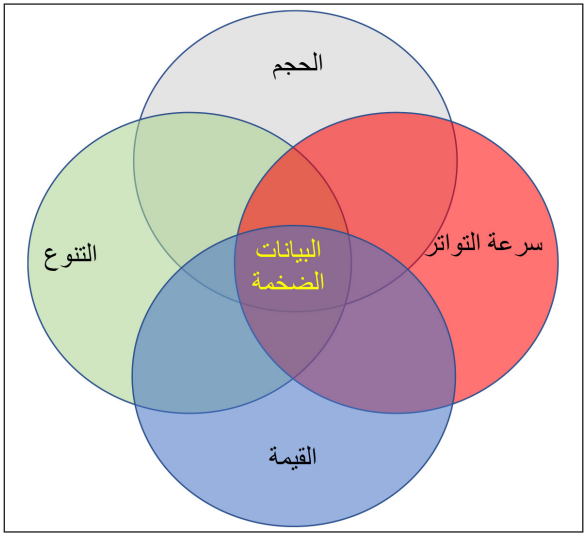
## ٢- ما هي البيانات الضخمة؟

تعرف البيانات الضخمة على أنها كميات كبيرة من البيانات التي يتم إنشاؤها/جمعها من التفاعلات اليومية مع المنتجات أو الخدمات الرقمية. وتحمل البيانات الضخمة صفة أو أكثر من الصفات التالية:

- كبيرة الحجم (Volume)
- سريعة التواتر (Velocity)
- عالية التنوع (Varity)
- لها قيمة كبيرة (Value)

فيما يعرف بالإنجليزية بـ Big Data 4 Vs.

شكل (١) خصائص البيانات الضخمة



وقد شهدت السنوات الأخيرة نموًا في كمية البيانات النصية الرقمية المتاحة، مما أدى إلى تكوين أفكار جديدة وبالتالي إدخال مجالات جديدة للبحث. وقد اجتذب تحليل تلك البيانات اهتمامًا كبيرًا، وأدى ذلك إلى حدوث تطورات سريعة في تقنيات تحليل البيانات الضخمة عبر مجموعة واسعة من التطبيقات، وشهدت الأوساط الأكاديمية وكذلك الصناعة تحول نحو المبادرات البحثية وقضايا البحث المعقدة بشكل متزايد.

وقد تم وصف البيانات الضخمة على أنها تطور معرفي من شأنه أن يغير الطريقة التي نعيش ونعمل ونفكر بها بسبب إمكاناتها. والهدف الأساسي من هذا التطور المعرفي هو الاستفادة من كميات هائلة من البيانات لتحويلها إلى معلومات تساعد على اتخاذ القرارات بصورة أفضل، وتعتمد القدرة على استخراج معرفة ذات قيمة من البيانات الضخمة على تحليل البيانات؛ وفقًا لـ Jagadish et al.

ومع إمكاناتها الهائلة لتطوير البحوث في مجالات العلوم الاجتماعية، تعمل البيانات الضخمة بشكل أساسي على تقليل الاعتماد على الممارسات العلمية الحالية واستبدال التقنيات التقليدية الأكثر تكلفة والتي تستغرق وقتًا طويلاً في جمع البيانات مثل الاستطلاعات أو المقابلات المتعمقة.

وتتنوع مصادر البيانات الضخمة لتشمل وسائل التواصل الاجتماعي، تطبيقات التليفون المحمول، شركات الاتصالات، تطبيقات تحديد المواقع GPS، المتاجر، البنوك، التسوق عبر الإنترنت وغيرها من المصادر التي تتيح بيانات منظمة أو غير منظمة.

وبغض النظر عن مصادر البيانات الرقمية، يثري هذا النوع الجديد من البيانات آفاق البحث ولديه القدرة على تطوير البحث في العلوم الاجتماعية والإنسانية من خلال ما يلي:

- قد تساعد أدوات جمع البيانات الضخمة المتقدمة، مثل تمشيط شبكة المعلومات (web scraping) والتقنيات التحليلية المبتكرة، مثل التعلم الآلي، في إنشاء منهجيات بحث جديدة،
- قد تكشف الأنواع الجديدة من البيانات عن أنماط جديدة للمجتمع البشري والسياسة والاقتصاد،
- قد تؤدي الأنواع الجديدة من البيانات إلى أنواع جديدة من الأسئلة البحثية التي تتجاوز منظور النظريات الراسخة.

وتتميز البيانات الضخمة بأنها قد تكون أرخص وأسرع من أساليب توفير البيانات الأخرى، كما أنها قد تكون مصدر يسهل الحصول على بيانات منه في المناطق أو الظروف التي يصعب جمع البيانات فيها بالأساليب التقليدية كما كان الحال في بداية انتشار كوفيد-١٩. بالإضافة إلى ذلك يمكن أن تستخدم البيانات الكبيرة من المصادر المختلفة ليس فقط لقياس انتشار الظاهرة، بل أيضاً لتفسير الظاهرة ومعرفة أسبابها والتنبؤ بتطورها.

## ٣- تجميع وتحليل البيانات الضخمة

تحليل البيانات الضخمة هو بحث كمي يتعامل مع كميات كبيرة من البيانات المنظمة أو غير المنظمة لحساب المؤشرات المختلفة وكشف الأنماط والارتباط بين الظواهر المختلفة. ويتضمن ذلك شقين رئيسيين هما المنهجيات المستخدمة في التحليل والأدوات المستخدمة في التحليل.

تتضمن البيانات الضخمة عادةً مجموعات بيانات بأحجام تفوق قدرة أدوات البرمجيات شائعة الاستخدام لالتقاط البيانات وإدارتها ومعالجتها في غضون فترة زمنية معقولة، لذا يعتمد تحليل البيانات الضخمة على أساليب وتقنيات وأدوات مختلفة مثل تلك الموجودة في تحليل النص ودراسات السوق وعرض البيانات والتحليل الإحصائي. يعتبر التعلم الآلي (ML) كجزء من الذكاء الاصطناعي مكونًا أساسيًا لتحليل البيانات، وهو أحد المحركات الرئيسية لثورة البيانات الضخمة. والسبب في ذلك هو القدرة على التعلم من البيانات وتقديم رؤى وقرارات وتوقعات مستندة على البيانات وتقديم تفسير لها، كما أنه يعتمد على الإحصاءات ويمكنه، على غرار التحليل الإحصائي، استخراج الاتجاهات من البيانات. وفقًا

لطبيعة البيانات المتاحة، فإن الفئتين الرئيسيتين لمهام التعلم الآلي هما: التعلم تحت الإشراف والتعلم دون إشراف. التعلم تحت الإشراف يتطلب معرفة كل من المدخلات والمخرجات المرغوبة، والهدف من هذه المهمة هو أن يتعلم النظام كيفية تحويل المدخلات من البيانات إلى مخرجات. الفئة الثانية هي التعلم دون إشراف حيث لا تكون المخرجات المرغوبة معروفة ويكتشف النظام نفسه النمط الذي تتبعه البيانات.

وبشكل عام، يمكن تقسيم البيانات إلى عدد من الفئات المختلفة، بما في ذلك البيانات الاجتماعية (مثل الإعجابات على تويتر وفيس بوك)، وبيانات حول التنقل والمواقع الجغرافية المكانية (مثل بيانات الاستشعار التي يتم جمعها من خلال الهواتف المحمولة أو صور الأقمار الصناعية)، والبيانات التي تم جمعها من الإدارة مصادر داخل الحكومة، ومجموعات بيانات نصية متعددة اللغات.

ويمكن أن تكون البيانات في شكل منظم أو غير منظم، وهناك حاجة لتحليل كل من البيانات المنظمة وغير المنظمة في التشغيل اليومي للمؤسسات لتحديد ردود أفعال العملاء، وتفضيل المنتجات للمستهلكين، وتخصيص المنتج، والمتطلبات التنظيمية الأخرى.

وتعد بيانات وسائل التواصل الاجتماعي من أهم مصادر البيانات الضخمة. ويتزايد إقبال المستخدمين لوسائل التواصل الاجتماعي على مشاركة أفكارهم وشكاواهم حول السياسة العامة أو للتعبير عن مواقفهم تجاه حدث معين، أو شخص، أو منظمة، أو منتج، .. إلخ. ويمكن أن يكون تحليل هذه البيانات مفيدًا جدًا للفهم واتخاذ القرار بشكل أفضل. وبالتالي، هناك حاجة لطرق وأدوات أكثر كفاءة يمكن أن تساعد في اكتشاف وتحليل المحتوى في الشبكات الاجتماعية عبر الإنترنت، لا سيما لأولئك الذين يستخدمون المحتوى الذي ينشئه المستخدم كمصدر للبيانات. علاوة على ذلك، هناك حاجة لاستخراج المزيد من المعلومات المفيدة والمخفية من العديد من المصادر عبر الإنترنت التي يتم تخزينها كنص ومكتوبة بلغة طبيعية داخل الشبكة الاجتماعية.

بشكل عام، تسمى عملية تحديد وفهم موقف الكاتب/المدون فيما يتعلق بموضوع ما في جزء من النص سواء كانت آراء "إيجابية" أو "سلبية" بتحليل المشاعر. تحليل المشاعر، كتطبيق حديث لوسائل التواصل الاجتماعي، هو تحليل حسابي للآراء والمشاعر والعواطف والمواقف نحو ظالظاهرة المعبر عنها بالكلمات. تُعرف مهمة تحديد واستخراج وتصنيف الآراء والمشاعر والمواقف المعبر عنها في المدخلات النصية

باسم تحليل المشاعر (يُعرف أيضًا باسم التنقيب عن الرأي أو مراجعة التنقيب أو استخلاص التقييم أو تحليل المواقف). يساعد هذا النوع من التحليل في تحقيق مجموعة متنوعة من الأهداف، بما في ذلك قياس الرأي العام حول الحركات السياسية، ودراسات السوق، وتقييم رضا العملاء، والتنبؤ بمبيعات الأفلام، من بين أمور أخرى. بمساعدة التعلم الآلي ومحتوى وسائل التواصل الاجتماعي -والتي تعتبر بيانات غير منظمة- يمكننا إنشاء نماذج قادرة على تحليل آراء الفئات المختلفة حول الموضوعات ذات الأهمية (الانتخابات، قضايا المرأة، القضايا السياسية أو الاقتصادية وما إلى ذلك) بالإضافة إلى ذلك، يمكن استخدام النماذج التي تم تطويرها لحساب مؤشرات من هذه البيانات. على سبيل المثال، يمكن تقسيم المدونين حسب آرائهم الإيجابية أو السلبية حول موضوعات معينة. يمكن أن تحليل البيانات على أساس الجنس أو العمر أو أي معايير أخرى، على سبيل المثال يمكن القول أن ٦٠٪ من النساء ضد موضوع معين، أو أن ٩٠٪ من الإناث التي تتراوح أعمارهن بين ١٥ و٢٩ عامًا لديهن انطباع إيجابي حول موضوع معين وما إلى ذلك.

تطبيق آخر مثير للاهتمام للتنقيب في وسائل التواصل الاجتماعي اكتسب شعبية في السنوات الأخيرة هو ما يسمى بنمذجة الموضوع. نمذجة الموضوع هي تقنية التنقيب عن النص لاكتشاف الموضوعات الشائعة في مجموعة من المستندات. من خلال تدريب نموذج موضوع على مجموعة من بيانات الوسائط الاجتماعية، سنتمكن من رؤية الكلمات التي تحدث بشكل متكرر معًا، والتي ستكون بمثابة موضوعات. عادةً ما يكون استخراج السمات المفيدة من بيانات نصية كبيرة يدويًا أمرًا غير ممكن. ومع ذلك، فإن تقنيات التنقيب عن النص باستخدام نمذجة الموضوع توفر آلية لاستنتاج الموضوعات تلقائيًا من مجموعة نصية.

ويمكن أن تستخدم نمذجة الموضوعات لاكتشاف الموضوعات المشتركة للمدونين. على سبيل المثال، استخدم الباحثون نمذجة الموضوعات حول كوفيد-١٩ لمعرفة الموضوعات الإخبارية المتعلقة بكوفيد-١٩ من وسائل الإعلام عبر الإنترنت في الصحف. أيضًا، استخدم الباحثون نمذجة الموضوع لفهم السياق السياسي لجنوب إفريقيا من خلال تحليل الآراء التي تمت مشاركتها على Twitter أثناء انتخابات الحكومة المحلية.

لذلك، فإن نمذجة الموضوع ضرورية لمعرفة الموضوعات التي تكون أخبارًا عن أحداث معينة (اجتماعية، سياسية، اقتصادية، إلخ ..).

## ٤- البيانات الضخمة من أجل التنمية

البيانات الضخمة من أجل التنمية هي تحديد مصادر البيانات الضخمة ذات الصلة بسياسات وتخطيط برامج التنمية، وذلك لتوفير تحليلات في وقت أقصر نسبيًا من الوقت الذي تحتاجه مصادر المعلومات الأخرى.

ومن أهم تطبيقات البيانات الضخمة في المجالات التنموية ما يلي:

- الإنذار المبكر: يساعد تحليل البيانات الضخمة من المصادر المختلفة في التنبؤ المبكر بالمشكلات التي قد تواجه أي مجتمع أو فئة مثل التنبؤ بارتفاع الأسعار ونقص المواد الغذائية في بعض المناطق وغيرها.
- صياغة السياسات: توفر البيانات الضخمة فرصة كبيرة للحصول على معلومات تساعد على وضع سياسات مستندة على الأدلة.
- التغذية المرتدة السريعة: يمكن لتحليل البيانات الضخمة التي تعكس الآراء حول السياسات والبرامج المختلفة أن يساعد في تعديل البرامج والسياسات.
- تقييم التدخلات: تقييم مدى نجاح التدخلات التي تقوم بها الدول، والتنبؤ بنتائج تلك التدخلات.
- فهم المجتمع محل الدراسة: الحصول على معلومات تساعد على فهم نمط حياة ورفاه المجتمع بصورة أفضل

وبالرغم من الأهمية التي اكتسبتها البيانات الضخمة في المجالات التنموية نظراً لتوافرها واتساع مدها والانخفاض النسبي في تكلفتها، إلا أن هناك اعتبارات منهجية وتحديات تتعلق بتغطية البيانات -سنعرضها بالتفصيل فيما بعد- تدعو إلى الدمج بين استخدام البيانات الضخمة إلى جانب الأساليب التقليدية في توفير المؤشرات والمعلومات التي يمكن الاعتماد عليها في بناء استراتيجيات وسياسات وقرارات مستندة على الأدلة.

## ٥- البيانات الضخمة وتمكين المرأة

تعد البيانات الضخمة أحد عوامل تمكين المرأة، ليس فقط لأنها تعد فرصة للحصول على بيانات بصورة سريعة لكن لأنها أيضاً تمثل فرصة لإشراك كافة الأطراف المعنية بقضايا تمكين المرأة، لا سيما النساء أنفسهن، في عملية اتخاذ القرار والتخطيط والمتابعة والتقييم.

فيما يلي أمثلة لاستخدام البيانات الضخمة في توفير مؤشرات حول المرأة المصرية:

| المقترح | استخدام قاعدة بيانات الرقم القومي في حساب مؤشرات الإنجاب[١] |
|---|---|
| الهدف | تقدم قاعدة بيانات الرقم القومي فرصة ذهبية لحساب عدد من المؤشرات لعل أهمها في مجال التمكين الاجتماعي للمرأة المؤشرات الخاصة بالإنجاب حيث يمكن حساب مؤشرات معدلات الإنجاب الكلي ومتابعتها بصورة سنوية وعلى مستوى الوحدات الإدارية الأصغر (المراكز، القرى، الأحياء)، بدلاً من الاعتماد على المسوح الأسرية التي يتم إجراؤها في سنوات متباعدة لا تسمح بسرعة التحرك من جانب متخذي القرار والتنفيذيين، كما لا يسمح تصميم وحجم عيناتها بحساب المؤشرات على الوحدات الإدارية الأصغر. |
| المقترح | استخدام بيانات شركات الاتصالات في حساب مؤشرات التنقل |
| الهدف | هناك ٤ شركات تقدم خدمات الهاتف المحمول في مصر، وتتوافر لدى هذه الشركات بيانات وقتية عن برج الاتصال الذي تلقى المستخدم المكالمات منه. هذه البيانات تعد فرصة كبيرة لدراسة تنقل المرأة خاصةً في أيام العمل خلال الأسبوع مما يعطي فرصة للتعرف على المسافات التي تقطعها المرأة يومياً، وتقدير الوقت والتكلفة الاقتصادية والاجتماعية للتنقل. كذلك يمكن من خلال هذه البيانات معرفة خطوط السير التي تستخدمها السيدات بكثافة مما يساعد على تطوير منظومة النقل وتوفير وسائل الانتقال الآمنة للمرأة. |
| المقترح | استخدام بيانات استخدام بطاقات الائتمان في عمليات الشراء |
| الهدف | تحليل بيانات الشراء باستخدام بطاقات الائتمان يمكن أن يساعد في تحليل نمط استهلاك الأسر المصرية والتنبؤ بالأزمات المتعلقة بالغذاء والخدمات الأساسية. |

١ قد ينظر إلى قاعدة بيانات الرقم القومي باعتبارها من البيانات الإدارية وليست من البيانات الضخمة. إلا أن حجم البيانات الضخم وسرعة تواترها مع تسجيل كل مولود وكل حالة وفاة بالإضافة إلى استخدامها في حساب مؤشرات هي ليست الهدف الأساس لقاعدة البيانات برشحها للتعامل معها بنفس تقنيات التعامل مع البيانات الضخمة.

| المقترح | استخدام تطبيقات تحديد الموقع في دراسة حصول المرأة على الخدمات |
|---|---|
| الهدف | يمكن من خلال هذه التطبيقات دراسة أنماط حياة السيدات بصفة عامة، فعلى سبيل المثال يمكن دراسة معدلات تردد السيدات على الأسواق وعلى أماكن تقديم الخدمات الحكومية وعلى مراكز الخدمات الصحية. دراسة معدل التردد على الخدمات الصحية على سبيل المثال يمكن ربطه بالمستوى الصحي للسيدات خاصةً. فيما يتعلق بالصحة الإنجابية، وهو ما يساعد بدوره على التنبؤ بمعدلات الإنجاب. |
| المقترح | تحليل بيانات فواتير الخدمات الأساسية |
| الهدف | يمكن استخدام بيانات الفواتير أو الشحن من قواعد بيانات شركات الكهرباء والمياه والغاز في دراسة انتظام الخدمات في محل السكن. كما يمكن استخدام هذه البيانات في تطوير مؤشرات الفقر ومستوى الرفاهة. |
| المقترح | استخدام محتوى منصات التواصل الاجتماعي في تقييم الخدمات التي تحصل عليها المرأة |
| الهدف | يمكن من خلال تحليل تدوينات الإناث على منصات التواصل الاجتماعي والتي تحتوي تعليقات على الخدمات المختلفة تقييم مستوى الخدمات التي تحصل عليها المرأة، والتعرف على الخدمات التي تعاني المرأة المصرية من عدم توافرها. وقد يمكن الاستفادة من ذلك بصورة أكبر إذا تم تحليل هذه التدوينات على مستوى المناطق وحسب الخصائص المختلفة للإناث اللاتي كتبن هذه التدوينات. |
| المقترح | استخدام محتوى منصات التواصل الاجتماعي في دراسة التوافق حول قوانين الأحوال الشخصية |
| الهدف | تطوير قوانين الأحوال الشخصية دائماً مثار جدل في مصر، وهو ما يؤدي إلى وجود آلاف التدوينات حول كل مشروع قانون يتم طرحه. تحليل هذه التدوينات بصورة مستمرة مع ظهور مشروعات القوانين الجديدة سيساعد على التعرف على رأي المواطنين في هذه المشروعات، وتضمين مخاوفهم ومقترحاتهم في المناقشات التي يقوم بها نواب البرلمان قبل التصويت على أي مشروع.<br>قد يكون هذا التحليل مفيد للتعرف بصورة سريعة على رد الفعل على التغييرات المتتالية سريعة الوتيرة المقترحة على مواد القوانين. |

| المقترح | دراسة العنف ضد المرأة |
|---|---|
| الهدف | يمكن من خلال دراسة التدوينات على مواقع التواصل الاجتماعي المختلفة حول حالات العنف ضد المرأة وربطها بعدد محاضر الشرطة الخاصة بالعنف ضد المرأة أو بنتائج المسوح التي تتم للتعرف على نسب السيدات التي تعرضن للعنف بناء نموذج إحصائي للتنبؤ بعدد حالات العنف ضد المرأة من خلال أعداد التدوينات حول نفس الموضوع. |
| المقترح | تحليل القيم الاجتماعية السائدة نحو تمكين المرأة |
| الهدف | تعكس التعليقات التي يدونها الأفراد على الأخبار والتدوينات المتعلقة بتمكين المرأة القيم الاجتماعية السائدة نحو تمكين المرأة في مجالات المختلفة. ويمكن من خلال تحليل محتوى هذه التدوينات متابعة التغير في قيم المجتمع نحو تمكين المرأة. |
| المقترح | التعرف على هموم المرأة المصرية من خلال دراسة الموضوعات التي يتم البحث عنها على محركات البحث المختلفة |
| الهدف | تمثل محركات البحث فرصة سانحة للتعرف على هموم واحتياجات المرأة المصرية التي قد تتغير من وقت لآخر خاصةً في أوقات الأزمات والمحن بصورة وقتية، مما يساعد الحكومة على الاستجابة بصورة أسرع لهذه الاحتياجات. |
| المقترح | دراسة الطلب على المرأة في سوق العمل من خلال طلبات التوظيف |
| الهدف | تعد قواعد بيانات طلبات التوظيف التي يمكن توفيرها من خلال مواقع التوظيف الإلكترونية وتطبيقات المحمول المشابهة مصدراً يمكن الاعتماد عليه في تحليل بيانات الطلب على العمالة في التخصصات المختلفة، ويمكن من خلالها التعرف على الوظائف التي يفضل أصحابها تعيين الذكور وتلك التي يفضل أصحابها تعيين الإناث وذلك من خلال تحليل المواصفات المطلوبة وما إذا كانت تشترط نوع معين، وكذلك من خلال تحليل خصائص من تم توظيفهم فعلاً عن طريق هذه المواقع. |
| المقترح | دراسة التحيز ضد المرأة في مكان العمل |
| الهدف | دراسة التدوينات التي تطلقها المشتغلات على مواقع التواصل الاجتماعي والتي تتحدث عن التحيز ضدهن في العمل يمكن أن تساعد على تحليل أنواع التحيز الذي تتعرض له المرأة في مكان العمل. |

| المقترح | تأثير التغيرات المناخية على المرأة |
|---|---|
| الهدف | تظهر تدوينات عديدة في الأوقات التي تزداد فيها آثار التغيرات المناخية تحكي عن المواقف التي تعرض لها أصحابها نتيجة لذلك، ويمكن من خلال تحليل هذه التدوينات التعرف على المشكلات والتحديات التي تواجه المرأة وآثارها عليها نتيجة التغير المناخي. |
| المقترح | دراسة توافر خدمات الصحة الإنجابية وبخاصةً وسائل تنظيم الأسرة |
| الهدف | يمكن من خلال تحليل التدوينات على وسائل التواصل الاجتماعي وعمليات البحث على محركات البحث المختلفة التعرف على احتياجات السيدات غير الملباة من الأنواع المختلفة من وسائل تنظيم الأسرة. |

## ٦- تحديات استخدام البيانات الضخمة في حساب مؤشرات تمكين المرأة

### ▪ اعتبارات الخصوصية

هناك اعتبارات متعلقة بالخصوصية يجب مراعاتها عند التعامل مع البيانات، فعلى سبيل المثال عند جمع بيانات المسوح يتم إخبار المستجيب للمسح بأهداف المسح والتأكيد على سرية البيانات الشخصية والحصول على موافقته قبل البدء في استيفاء استمارة المسح معه. أما في حالة البيانات الكبيرة، فإنه رغم أن الشخص قد يكون أدلى بالبيانات أو المعلومات بصورة طوعية إلا أنه في الغالب قد أدلى بها لأسباب محددة وليس لاستخدامها في أغراض الدراسة التي سيتم استخدام البيانات الكبيرة في إطارها.

بالإضافة إلى ذلك، عند استخدام ما يتم تدوينه على مواقع التواصل الاجتماعي من خلال الحسابات الشخصية التي تظهر شخصياتهم وخصائصهم المختلفة وربما ظروف معيشتهم فإن ذلك قد يعرض المدونين للخطر خاصةً في حالة ربط ما يدونونه ببياناتهم التعريفية في ملفات البيانات المستخدمة. وبالرغم من أن هناك العديد من مبادئ الخصوصية التي أعتمدتها الكثير من المنظمات الدولية في التعامل مع البيانات إلا أن بعض النقاط لم يتم التوصل لاتفاق بشأنها خاصةً فيما يتعلق بالحد الذي تصبح عنده البيانات تعريفية أو غير تعريفية للمدون.

### ▪ تمثيل المجتمع

عـادةً يتـم جمـع البيانـات مـن خـلال المسـوح مـن عينـة ممثلـة للمجتمـع مـما يسـمح بتعميـم نتائـج هـذه المسـوح عـلى المجتمـع ككل، وهـو مـا قـد لا يتحقـق عنـد اسـتخدام البيانـات الضخمـة. ويتوقـف مـدى تمثيـل البيانـات الضخمـة للمجتمـع عـلى مصـادر هـذه البيانـات، فعـلى سـبيل المثـال البيانـات التـي يتـم الحصـول عليهـا مـن قواعـد البيانـات الإداريـة مثـل قاعـدة بيانـات الرقـم القومـي أو قواعـد بيانـات شركات الكهربـاء والميـاه هـي بيانـات ممثلـة للمجتمـع بينـما البيانـات التـي يتـم الحصـول عليهـا مـن وسـائل التواصـل الاجتماعـي غـير ممثـل للمجتمـع لعـدة أسـباب نوجـز أهمهـا فيـما يـلي:

- المدونـون عـلى وسـائل التواصـل الاجتماعـي هـم مـن يسـتطيعون القـراءة والكتابـة ولديهـم إمكانيـة الوصـول للإنترنـت ولديهـم حسـابات عـلى مواقـع التواصـل الاجتماعـي، وهـو مـا يعنـي أن البيانـات أو الآراء التـي يتـم تجميعهـا مـن هـذه المواقـع لا تمثـل الأميـين أومـن ليـس لديهـم إمكانيـة الوصـول للإنترنـت أو مـن لا يمتلكـون حسـابات عـلى هـذه المواقـع.
- أسـاليب تجميـع تدوينـات أصحـاب الحسـابات عـلى منصـات التواصـل الاجتماعـي لا تضمن تجميـع كل ما كتـب عـن الموضـوع محـل الدراسـة كما أنهـا لا تضمن أن العينـة التي تـم تجميعها من هـذه التدوينات هـي عينـة غـير متحيـزة لـرأي معـين، بالإضافـة إلى ذلـك هـذه التدوينـات هـي تدوينـات مـن تطوعـوا للكتابـة عـن الموضـوع وبالتـالي لا تعكـس رأي مـن لم يقومـوا بالتدويـن حـول الموضـوع والذيـن في كثـير مـن الأحيـان يكـون لهـم آراء مخالفـة. عـلى سـبيل المثـال إذا تـم تجميع مـا يكتب حول تقييـم الخدمات الصحيـة، نجـد أن مـن يقومـون بالتدويـن هـم عـادةً مـن واجهـوا مشـكلة في الحصـول عـلى الخدمـة أو واجهـوا مشـكلة في جـودة الخدمـة، لكـن مـن لديهـم رضـا عـن الخدمـة لا يقومـون بالتدويـن عنهـا.

### ▪ الموارد البشرية والمالية

كـما سـبق أن ذكرنـا، يتطلـب تجميـع وتحليـل البيانـات الضخمـة تقنيـات عاليـة واسـتخدام برامـج حاسـوب ذات قـدرات عاليـة لتحليـل هـذا الكـم مـن البيانـات، مـما يتطلـب توفـير وتدريـب كـوادر بشريـة قـادرة عـلى التعامـل مـع هـذه البرامـج وهـو مـا قـد لا يتوافر بسـهولة نتيجـة نـدرة الكـوادر البشريـة المتخصصة في هـذا المجـال في مـصر. كـما أن توفـير هـذه البرامـج وتوفـير أجهـزة حاسـوب ذات مواصفـات خاصـة يمكنهـا

التعامل مع هذه البرامج وهذا الكم من البيانات الضخمة يتطلب موارد مالية كبيرة. غير أن توفير هذه الموارد في البداية سيضمن استخدامها في مجالات تنموية مختلفة مما يوفر في المستقبل كثير من الموارد المالية التي يتم إنفاقها لجمع البيانات بالطرق التقليدية.

- **اللغات المحلية والعامية**

يستخدم معظم المدونين على مواقع التواصل الاجتماعي اللغات المحلية والعامية في تدويناتهم، وهو ما يجعل تحليل الكلمات المستخدمة بواسطة برامج تحليل المحتوى صعباً، إذ قد يكون لنفس الكلمة معانٍ مختلفة في اللغات المحلية المختلفة، كما أن المدونين قد يستخدمون كلمات بصورة هزلية للتعبير عن عكس ما تدل عليه الكلمة. بالإضافة إلى ذلك مازالت برامج تحليل المحتوى باللغة العربية تحتاج لكثير من التطوير.

- **صعوبة حساب المؤشرات المتعارف عليها**

قد لا يمكن من خلال البيانات الضخمة توفير البيانات المطلوبة لحساب المؤشرات المتعارف عليها بصورة مباشرة، وهو ما قد يتطلب في هذه الحالة وضع منهجيات لمؤشرات بديلة يمكن حسابها من خلال البيانات المتاحة وتقدم تقدير مقبول للمؤشر الأصلي الذي لا يمكن حسابه من البيانات المتاحة.

## ٧- التوصيات

بالرغم من التحديات التي تواجه البيانات الضخمة كمصدر للبيانات لحساب المؤشرات الخاصة بتمكين المرأة إلا أنها مازالت تشكل فرصة كبيرة للحصول على البيانات اللازمة بصورة مستمرة وبتكلفة أقل، كما تسمح بالحصول على مؤشرات على المستوى المحلي. وتحتاج مصر إلى مزيد من البيانات التي تسمح بمتابعة وتقييم أهداف التنمية المستدامة على المستوى المحلي، والاستراتيجيات التنموية لا سيما الاستراتيجية الوطنية لتمكين المرأة. لذا يعد الاهتمام بترسيخ استخدام البيانات الضخمة في توفير المؤشرات التنموية بصورة عامة ومؤشرات تمكين المرأة بصورة خاصة فرصة سانحة لتطوير آليات المتابعة والتقييم. ويتطلب ذلك ما يلي:

- تحديد المؤشرات التنموية التي لا يمكن توفيرها من المصادر التقليدية المتاحة، ووضع المنهجيات المناسبة لتقديرها من المصادر المختلفة للبيانات الضخمة.
- إتاحة استخدام مصادر البيانات الضخمة المختلفة بعد وضع الشروط التي تضمن سرية البيانات الشخصية، على سبيل المثال قاعدة بيانات الرقم القومي، قواعد بيانات المعاملات البنكية وعمليات الشراء باستخدام الكروت البنكية.
- توفير الموارد المالية بمساعدة شركاء التنمية المختلفين للتوسع في استخدام البيانات الضخمة في المجالات التنموية وما يشمله ذلك من تعيين خبراء في هذا المجال لوضع المنهجيات اللازمة لتقدير المؤشرات التنموية المختلفة من البيانات الضخمة وتطبيق هذه المنهجيات ودراسة نتائج هذا التطبيق، ووضع النماذج الإحصائية التي تساعد على استخدام البيانات الضخمة في التنبؤ بسلوك المجتمع ككل رغم عدم تمثيل البيانات الضخمة له.

تعد ندرة الكوادر البشرية في هذا المجال من التحديات التي تواجهها مصر. وهو ما يتطلب مايلي:

- التوسع في استحداث التخصصات العلمية المتعلقة بالبيانات الضخمة في الجامعات المصرية.
- ارسال بعثات للخارج من خريجي الجامعات المصرية المتخصصين في تكنولوجيا المعلومات وتحليل البيانات لدراسة منهجيات وتطبيقات البيانات الضخمة في المجالات التنموية.

تعد مسألة الحفاظ على خصوصية الأفراد من التحديات الحاكمة لاستخدام البيانات الضخمة، والتي قد تؤدي إلى محدودية القدرة على استخدامها. لذا يتطلب التوسع في هذا الاستخدام ما يلي:

- وضع وتطوير مواثيق أخلاقية تضمن سرية البيانات التعريفية والشخصية للمدونين أو أصحاب البيانات بما يحميهم من أي إيذاء قد يتعرضون له نتيجة استخدام بياناتهم.
- تدريب المتعاملين مع هذه البيانات على الحفاظ على خصوصية وسرية البيانات الشخصية، ووضع ضوابط للمساءلة والمحاسبة في حالة الإخلال بالمواثيق الأخلاقية المتعلقة بذلك.

## ٨- المصادر


- H. V. Jagadish et al., "Big data and its technical challenges", Commun. ACM, vol. 57, no. 7, pp. 86-94, 2014
- Ledwaba, M. and Marivate, V., 2022. Semi-supervised learning approaches for predicting South African political sentiment for local government elections. arXiv preprint arXiv:2205.02223.
- R. P. Fauzie Afidh and Z. A. Hasibuan, "Indonesia's News Topic Discussion about Covid-19 Outbreak using Latent Dirichlet Allocation," 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020, pp. 1-6, doi: 10.1109/ICIC50835.2020.9288596.
- M. A. Beyer and D. Laney, "The importance of ‘big data’: A definition", 2012.
- UN Women, Gender equality and big data, 2018
- V. Mayer-Schönberger and K. Cukier, Big Data: A Revolution That Will Transform How We Live Work and Think, Boston, MA:Houghton Mifflin Harcourt, 2013.

# استخدام البيانات الضخمة في توفير مؤشرات حول تمكين المرأة