

علا نبيل القصاص . مشروعات البيانات الرقمية Big Data في المكتبات : دراسة استكشافية / علا نبيل القصاص ؛ إشراف أمنية مصطفى صادق ، نوره ذكي الشنواني . - المنوفية : ع.ن. القصاص ، ٢٠١٩ . - رسالة ماجستير ، كلية الآداب ، جامعة المنوفية .

عرض

## علا نبيل القصاص

قسم المكتبات والمعلومات  
كلية الآداب - جامعة المنوفية

### تمهيد

تختلف طبيعة البيانات الضخمة بشكل كبير عن البيانات الأخرى ، فليست كل البيانات مهيكلة ويمكن وضعها في جداول وإدراجها في برامج التحليل كالأرقام ، ولكن يوجد بيانات أخرى غير مهيكلة كالصور ، والمقاطع الصوتية ، والفيديو ، ورسائل البريد الإلكتروني ، وتغريدات تويتر ... إلخ وبالطبع تلك البيانات غير المهيكلة تمثل نسبة أكبر من البيانات المهيكلة حيث يتم إنتاجها كل دقيقة يومياً ؛ وبالتالي لا يمكن تنسيقها في شكل جدولي ، الأمر الذي يتطلب وجود أدوات وتقنيات تناسب هذا النوع من البيانات لتحليلها والخروج بنتائج مثمرة .

وتعتمد تكنولوجيا البيانات الضخمة Big Data على ثلاث مراحل أساسية أولها تجميع البيانات Data Collection ، وتحليلها Data Analysis ، ومن ثم التصور والتطبيق Visualization and Application ، وهذه المرحلة الأخيرة هي المرحلة التي تظهر بها نتيجة تحليل البيانات وتتاح فيها البيانات المحللة للمستخدمين بشكل مفسر و متكامل ، ثم تُستخدم في نهاية المطاف لتوجيه وصنع القرار .

ومع استمرار تطور البيانات الضخمة وتقنيات الحوسبة السحابية، سيتحول التركيز من كيفية تخزين ومعالجة كميات كبيرة من البيانات إلى كيفية اتخاذ قرارات مدروسة ودقيقة وفعالة و زهيدة التكلفة ، وقد يتطلب ذلك إحداث نقلة نوعية في كيفية إدارة البيانات و تحليلها.

بناءً على ذلك جاءت فكرة الرسالة وهي استكشاف استخدام البيانات الضخمة في إقامة مشروعات بالمكتبات .

### مشكلة الدراسة

يمر العالم اليوم بالكثير من الطفرات العلمية والتكنولوجية في شتى قطاعات الحياة ، مما أدى إلى انفجار المعلومات بشكل يصعب السيطرة عليه والتعامل معه بالطرق التقليدية ، ولعل قطاع تكنولوجيا المعلومات أحد القطاعات التي شهدت تطورات تقنية متلاحقة خاصة فيما يتعلق بمجال إدارة البيانات والمعلومات ، وسبل تنظيمها وتخزينها في ظل تزايد وتراكم البيانات وتعدد أشكالها وصيغها و أنماطها.

■ وتواجه المكتبات ظاهرة نمو غير مسبوق للبيانات والذي عرف حالياً بظاهرة البيانات الضخمة ، وذلك نتيجة لانتشار الأجهزة المتنقلة والمشاركة على منصات التواصل الاجتماعي ، وبذلك تصبح المكتبات عاجزة عن إدارة تلك البيانات بالطرق التقليدية وبالتالي عدم القدرة على إتخاذ القرارات المناسبة في حالة عدم توفر التقنيات الملائمة لإدارتها وتحليلها بشكل فعال لتحقيق أقصى إفادة منها .

■ ومن شأن تلك البيانات الضخمة أن تساعد في التوصل إلى فهم أكثر شمولاً وتفصيلاً لاحتياجات وخصائص المستخدمين بالمكتبة ، وذلك عن طريق جمع وتخزين مجموعة من البيانات المتنوعة

الخاصة بهم وربطها بالبيانات الأخرى داخل المكتبة ، وتحليلها ومن ثمّ تصورها بشكل يُسهل إستيعابها ، وبناء على ذلك يتم إتخاذ القرارات المثمرة وإيضاح الكثير من الرؤى الخفية ، وينتج عن ذلك التنبؤ بالتغيرات التي تطرأ على الرغبات المفضلة والسلوكيات بين مجموعات محددة من المستفيدين بسهولة وبالتالي إرتفاع مستوى الرضا بين المستفيدين بالمكتبات .

- وبناءً عليه تتمثل مشكلة الدراسة في عدم وجود دراسة أكاديمية عربية تقوم باستكشاف وضع البيانات الضخمة وتطبيقها في مجال المكتبات وأوجه الاستفادة من ذلك عن طريق إقامة ما يسمى بمشروعات البيانات الضخمة Big Data Projects ومدى تأثير ذلك على سوق العمل .

### أهمية الدراسة

- تتبع أهمية الدراسة من أهمية البيانات الضخمة وضرورة الاهتمام بتحليلها لاستخراج معلومات ذات قيمة يُعتمد عليها فيما بعد لاتخاذ قرارات مصيرية، وبناء على ذلك تناولت الدراسة التعريف بماهية البيانات الضخمة وفرص وتحديات تطبيقها في المكتبات ، و الأدوات والتقنيات المستخدمة في استخراج وتحليل تلك البيانات الضخمة و الاستفادة منها ، وإمكانية إنشاء مشروعات قائمة على البيانات الضخمة في المكتبات ، كما أنها ساعدت على توضيح ماهية تقنية التصور Visualization وكيفية الاستفادة منها في توضيح نتائج المشروعات داخل المكتبات باعتبارها مرحلة أساسية من مراحل البيانات الضخمة ، وفي النهاية مدى تأثير تلك التكنولوجيا الحديثة في سوق العمل و أهم المسميات الوظيفية المستحدثة المتطلب توافرها للتعامل مع البيانات الضخمة .
- يمكن الاستفادة من نتائج هذه الدراسة الاستكشافية للخروج بموضوعات بحثية جديدة ومحاولة توظيفها في مجال المكتبات والمعلومات .

### أهداف الدراسة

إنطلاقاً من مشكلة الدراسة سعت الدراسة إلى تحقيق ما يلي:

- التعرف على ماهية البيانات الضخمة ونشأتها وطبيعتها وأهميتها.
- دراسة وتحديد الأساليب والأدوات والتقنيات المستخدمة في تجميع البيانات الضخمة وتحليلها وكذلك طرق عرضها.
- دراسة التصور Visualization كأهم مراحل و أدوات عرض البيانات الضخمة
- تحليل مكونات أو هيكله مشروعات مختلفة للبيانات الضخمة في المكتبات
- استنباط مجالات جديدة في سوق العمل تعتمد على البيانات الضخمة وتقنياتها ، وذلك من خلال التعرف على وظائف فريق العمل بالمشروعات .
- استكشاف مدى ارتباط البيانات الضخمة بمقررات علوم المكتبات التي يتم تدريسها وكيفية تدريسها لأجيال أخصائي المكتبات .

**مجتمع وعينة الدراسة**

يشمل مجتمع الدراسة المشروعات المطبقة للبيانات الضخمة بشكل عام وهي كالتالي :

**جدول يوضح مجتمع الدراسة**

Open university Library Data Project	<a href="http://www.open.ac.uk/blogs/LibraryData">/http://www.open.ac.uk/blogs/LibraryData</a>
Ag Data Commons Beta : National Agricultural Lib	<a href="https://data.nal.usda.gov">https://data.nal.usda.gov</a>
Library Analytics Toolkit : Harvard university	<a href="https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit">https://osc.hul.harvard.edu/liblab/projects/library-analytics-toolkit</a>
Big UK Domain Data for the Arts and Humanities	<a href="https://buddah.projects.history.ac.uk">https://buddah.projects.history.ac.uk</a>
Library Impact data Project	<a href="https://library.hud.ac.uk/blogs/lidp">https://library.hud.ac.uk/blogs/lidp</a>

**فصول الدراسة**

- **الإطار المنهجي للدراسة** : تم عرض منهجي للدراسة ، وذلك من خلال العناصر التالية : تمهيد ، مشكلة الدراسة ، أهمية الدراسة ، أهداف الدراسة ، تساؤلات الدراسة ، حدود الدراسة ، منهج الدراسة ، أدوات جمع البيانات ، الدراسات السابقة وفصول الدراسة .
- **الفصل الأول :- البيانات الضخمة : الماهية و الأهمية**

تناولت الباحثة فيه تاريخ وتطور البيانات الضخمة ، والتعريفات التي تم الوصول إليها وخصائص و أهمية البيانات الضخمة ، والتكنولوجيات ذات الصلة بها ومراحل التعامل معها وتحديات وفرص استغلال تلك التقنية وإطار تطبيقها بشكل عام وفي المكتبات بشكل خاص وختاماً بمجالات الاستفادة من البيانات الضخمة بشكل عام وفي المكتبات والمعلومات بشكل خاص .

- **الفصل الثاني :- تقنيات و أدوات البيانات الضخمة في المشروعات : التخزين والتحليل والتصور**

وفيه تحدثت الباحثة عن طبيعة البيانات الضخمة ، ثم عرضت الأدوات والتقنيات الملائمة لتخزين وتحليل تلك البيانات عن طريق دراسة تكنولوجيا Hadoop و تكنولوجيا NOSQL ومن ثم عقد مقارنة فيما بينها ، ثم دراسة مجال التصور Visualization والغرض منه و أهم الأدوات المفتوحة المصدر لتمثيل تلك البيانات الضخمة مثل Tableau وغيرها من الأدوات الأخرى ، كما تم دراسة الوسائل الخاصة بعرض تلك البيانات وطبيعة كل وسيلة منها حسب نوع البيان المراد تمثيلة .

- **الفصل الثالث :- مشروعات البيانات الضخمة في المكتبات : دراسة وتحليل وإستنتاج**

تناولت الباحثة خلال ذلك الفصل النموذج الخاص بتطبيق مشروعات البيانات الضخمة في المؤسسات المختلفة والمنهجية المتبعة في ذلك وعناصر نجاح تلك المشروعات الضخمة ، ومن ثم

إعداد قائمة المراجعة الخاصة بتحليل الثلاث مشروعات محل الدراسة وهما Open University Ag Data و Library Analytics Toolkit : Harvard University ، Library Data Project Commons ومن ثم إجراء مقارنة بين الثلاث مشروعات .

### ■ الفصل الرابع : مجالات الاستفادة من البيانات الضخمة في سوق العمل

وفيه تم استعراض أمثلة فعلية حول مجالات الاستفادة من البيانات الضخمة من قبل شركات ومؤسسات ضخمة ودور المكتبات في الاستفادة من هذا الوعاء المعلوماتي الجديد، ثم عرض فريق عمل مشروعات البيانات الضخمة و الأشخاص والمهارات المطلوب وجودها للتعرف على المسميات الوظيفية الجديدة ومدى تأثير ذلك على مهنة أخصائي المكتبات والمعلومات وكيفية تطوير مهاراته للتماشي مع تلك التكنولوجيا، ومن ثم عرض بعض المواقع والمؤسسات التعليمية التدريبية في مجالات علوم البيانات بشكل عام والبيانات الضخمة بشكل خاص للتعرف على البرامج التعليمية المؤهلة للطلاب في سوق العمل وكيفية توظيف تلك البرامج التعليمية في علوم المكتبات والمعلومات .

### أهم نتائج الدراسة:-

لقد توصلت الباحثة إلى عدة نتائج، فيما يلي أبرزها :

١. البيانات الضخمة Big Data في أصل الموضوع بيانات عادية ، ولكنها تمتاز بضخامة الحجم Volume نتيجة حجم المستخدمين لها على مستوى العالم ، وتمتاز أيضا بالسرعة Velocity من حيث توليد وتحديث هذه البيانات ، وأيضاً بتنوع الهيئة والشكل Format الذي تأتي عليه البيانات ، وبالتالي تحتاج إلى تقنية أكثر تطوراً في التخزين والمعالجة ، ومن خلال المعالجة المناسبة يمكن تحويل هذه البيانات من صورتها الخام إلى قيمة معرفية Value وهو جوهر الاستفادة .
٢. التحول من التركيز على المنتج إلى التركيز على المستفيد و احتياجاته ، حيث سمحت البيانات الضخمة Big Data بالتعرف على الأنماط الإستهلاكية ، واحتياجات الشرائح المختلفة من المستهلكين؛ وبالتالي إمكانية تصميم خدمات تراعي هذه الاختلافات ، وقد يتطور الأمر إلى تصميم المنتج أو الخدمة لاحتياجات عميل واحد .
٣. من أفضل تقنيات التعامل مع البيانات الضخمة Hadoop الذي يتميز بسرعة معالجة البيانات في أماكن تواجدتها في وقت واحد ، وذلك بسبب اتباعه طريقة التوزيع والتقسيم للبيانات في عدة أجهزة ، حيث يتم معالجتها معاً .
٤. يوجد أدوات وتقنيات خاصة بتصوير البيانات بسيطة للغاية للمبتدئين مثل Data Hero حيث لا تحتاج إلى قدرات تقنية عالية ، كما يوجد أدوات معقدة في التعامل حيث تقوم بعمليات فرز وتجميع وتصنيف البيانات مثل Gephi و Weka ؛ وتلك تناسب فئة علماء البيانات Data Scientists .
٥. يوجد عدة أشكال مختلفة لعرض البيانات الضخمة ويتم تحديد الشكل المناسب لعرض بيانٍ ما ليتم تصوره بشكل فعال ، على سبيل المثال في حالة تمثيل منهجية انتشار البيانات ومدى ارتباط نقاط البيانات في مجالٍ ما فأنسب شكل لتمثيل ذلك هو Box Plots وبالتالي لا يمكن الاستفادة من البيانات الضخمة دون تصورها .

٦. هناك مسميات وظيفية مستحدثة لأمين المكتبة أيضا في إدارة البيانات الضخمة مثل **باحث أعمال Business Researcher** و **محلل البيانات الوصفية Metadata Analyst** وغيرها، وذلك يؤكد على مدة تأثير هذا المجال التكنولوجي الضخم " البيانات الضخمة " في سوق العمل وضرورة تعليمه والتدريب عليه من خلال تلك البرامج للتمكن من الاستفادة منه .

### أهم توصيات الدراسة :

بناءً على نتائج الدراسة، توصلت الباحثة إلى عدة توصيات، فيما يلي أبرزها :

- ١- ضرورة إضطلاع الجمعيات والاتحادات المهنية بإعادة صياغة التوصيف الوظيفي لإختصاصي المكتبات والمعلومات للتوافق مع علم البيانات Data Science
- ٢- الاستفادة من شبكات التواصل الاجتماعي كأداة لجمع البيانات الضخمة من أجل تحليلها والاستفادة منها في تحسين واقع المكتبات .
- ٣- اضطلاع أقسام المكتبات والمعلومات باستحداث برامج تعليمية سواء في مرحلة الليسانس أو مرحلة الدراسات العليا، بحيث توفر للدارس الحصول على مهارات تحليلية في مجال البيانات الضخمة تمكنه من التعامل معها وتوظيفها في قطاع المكتبات .
- ٤- اتخاذ مشروعات البيانات الرقمية الضخمة في المكتبات محل الدراسة بمثابة منهجية يمكن الاعتماد عليها من قبل مكتباتنا العربية .
- ٥- الاهتمام بإعداد البحوث الأكاديمية حول القضايا المتعلقة بالبيانات الضخمة في المكتبات ، وزيادة الوعي حول هذا المجال الجديد عن طريق إعداد الندوات وورش العمل والمؤتمرات العلمية التي تناقش سبل الإفادة من البيانات الضخمة.