اسهامات تنقيب البيانات في تنمية القطاعات الاقتصادية بين الفرص والتحديات

The contribution of data mining to the development of economic sectors opportunities & challenges

عبد الله بن الضب [1]، عبد الجليل طواهير [2]★

[1] المركز الجامعي ايليزي (الجزائر)، bendob.abdallah@cuillizi.dz

[2] جامعة قاصدي مرباح-ورقلة (الجزائر)، touahir.abdeldjalil@gmail.com

تاريخ الإستلام: .../.../..... تاريخ قبول النشر: .../.../..... تاريخ النشر: ../.../.....

ملخص:

تهدف الورقة البحثية إلى أبرز أهمية التنقيب في البيانات باعتباره المكون الأول للبيانات الضخمة من جهة، وكمدخل اداري حديث يهتم باستكشاف المعرفة في قواعد البيانات والتي أصبحت من الاهتمامات الكبيرة لأغلب القطاعات الاقتصادية على مستوى الدول الأمر الذي يساعد على بناء وتحليل التنبؤات المستقبلية في مجال اتخاذ القرار. من أهم نتائج هذه الدراسة بيان المراحل المختلفة للتنقيب عن البيانات الضخمة وتحديد أبرز التحديات؛ حيث يبقى التنقيب عن البيانات الضخمة معقدا في ظل البيانات غير المتجانسة مع صعوبة التوفيق بين ثنائية جودة البيانات والحفاظ على الخصوصية.

الكلمات مفتاحية: تنقيب البيانات؛ بيانات ضخمة؛ تنمية؛ قطاعات اقتصادية.

Abstract:

This research aims to show the importance of data mining as the first component of big data, In addition, it is a modern administrative entrance interested in exploring knowledge in databases on the other hand, which has become of great interest to most economic sectors at the international level, which helps to build and analyze future forecasts in the field of decision-making. One of the most important results of this study is to show the different stages of big data mining and to identify the most important challenges; Big data mining remains complex in light of heterogeneous data, with the difficulty of balancing data quality and maintaining privacy.

Keywords: data mining; big data; development; economic sectors.

★**المؤلف المرسل**. touahir.abdeldjalil@gmail.com

1. مقدمة :

يعد تضخم المعلومات أحد أبرز أشكال التقدم الذي تمر به البشرية في مرحلة تفوق العلوم والتقنيات، حيث لا يخفى على أحد الأثر الذي تركه هذا التفوق في حياة الإنسان، ومما لا شك فيه أن المجتمعات التي استملكت المعلومة والمعرفة هي الآن التي تحكم العالم وتتصدر مشهد الدول المتقدمة والعصرية.

حيث أصبحت مخازن البيانات وقواعد البيانات التي تعتمد عليها أغلب النظم والتطبيقات، أصبحت ممتلئة بكم كبير وضخم من البيانات العشوائية والتاريخية، ونستطيع القول فعلاً أن البيانات لا تنام ولا تتوقف عن النمو. مع هذا النمو في البيانات، أصبحنا أمام تحد كبير يتمثل في كيفية الاستفادة من هذه البيانات المخزنة في مستودعاتها، وكيفية القيام بعملية استخراج المعلومة المفيدة منها، وهذا التحدي أصبح لا يمكن حله بواسطة الطرق التقليدية لتحليل البيانات، لذلك فإننا بحاجة إلى تقنيات غير تقليدية تستوعب التعامل مع الكميات الضخمة للبيانات بالإضافة للقدرة على التعامل مع الأنواع والاشكال المختلفة للبيانات، وهذا يعني تقنيات على مستوى عالي من الكفاءة والذكاء، فمن المؤكد أن التنقيب في البيانات Data mining هو من الموضوعات بالغة الأهمية ... وإلا ما الذي يجعلها أحد الموضوعات الأكثر أهمية لرئيس الولايات المتحدة أوباما Obama ؟ فالمتتبعون لنشاط رئيس الدولة الأكثر تحكماً في العالم سيعرفون أنه بتاريخ 17 ديسمبر 2013، كان في لقاء مع قيادات تقنية المعلومات والذين كان منهم Tim Cook من شركة أبل Apple و Eric Schmidt من جوجل وكذلك المدراء التنفيذيين لكل منTwitter ، Microsoft، Facebook، Salesforce، Netflix، Etsy، Dropbox، !Yahoo، Zynga، Sherpa Global، Comcast، !!!LinkedInوذلك لمناقشة موضوع التنقيب في البيانات؛ كما أنه في خطاب له بتاريخ 17 يناير 2014 عاود من جديد طرح الموضوع من خلال الدعوة إلى إجراء إصلاحات في نظام التنقيب عن البيانات (حايك، 2014)

انطلاقا مما سبق تحاول الورقة معالجة الإشكالية

كيف يساهم تنقيب البيانات في ظل الفرص التي يوفرها في دعم القطاعات الاقتصادية؟

ومن أجل معالجة الإشكالية نطرح الأسئلة الفرعية الآتية

أولا: ما هو مفهوم تنقيب البيانات في ظل الأدب النظري لظهور هذا المفهوم؟

ثانيا: ما هي مراحل التنقيب في البيانات؟

ثالثا: ما مدى ضرورة عملية تنقيب البيانات لمختلف مكونات البيانات الضخمة؟

رابعا: ماهي مهام تنقيب البيانات؟

خامسا ما مدى تطبيق تنقيب البيانات الضخمة في مختلف مجالات القطاعية؟

سادسا ماهي التحديات التي تواجه تنقيب البيانات؟

2. الخلفية النظرية لمفهوم التنقيب البيانات:

1.2 ماهية تنقيب البيانات

أدى الانتشار الواسع لتكنولوجيا المعلومات وسهولة حصولهم على زيادة استباقية في حجم المعلومات التي لم يتم ملاحظتها من قبل في التاريخ، مما يجعل مسألة البيانات الكبيرة على الإنترنت الموضوع الأكثر جدلا من حيث جدوى وجوده بهذه الطريقة العشوائية. في منتصف التسعينيات ظهر مصطلح تنقيب البيانات في الولايات المتحدة الأمريكية (قرزيز، 2021، صفحة 200) ، من هناك أصبح ما يسمى بتعدين أو تنقيب البيانات تقنية لاستخراج المعرفة من كميات هائلة من البيانات، بناءً على الخوارزميات الرياضية، والتي هي أساس التنقيب عن البيانات وهي مشتقة من العديد من العلوم مثل الإحصاء والرياضيات والمنطق وعلوم التعلم والذكاء الاصطناعي والأنظمة الخبيرة والعلوم. التعرف على الأنماط وعلوم الآلة. وغيرها من العلوم التي تعتبر علوم ذكية غير تقليدية (أبرار، 2021).

يستخدم هذا الشكل من البيانات الضخمة طرقا فاعلة من تحليلها من أجل استخراج المعلومات المفيدة، ومن ثم اكتشاف الأنماط الخفية البيانات، مثلما هو الحال في دراسة احتياجات وحالات العملاء من خلال التسوق؛ وتقديم المقترحات مثل موقع أمازون للتسوق (سعد مأمون و أحمد عثمان، 2015)

يُطلق عليه أحيانًا اسم ''اكتشاف المعرفة''، ولم يُصاغ مصطلح ''التنقيب عن البيانات'' حتى التسعينيات. لكن أساسه يتشابك في العديد من التخصصات العلمية: الإحصاء (الدراسة العددية لعلاقات البيانات)، الذكاء الاصطناعي (ذكاء شبيه بالبشر تعرضه الآلات)، والتعلم الآلي (الخوارزميات التي تتعلم من البيانات لعمل تنبؤات) والمجال المعرفي للأعمال. نتيجة للنمو في هذا المجال، (طيار أحسن، شلابي عمار، بلا تاريخ) .

بدأت العديد من الشركات الكبيرة العمل معًا في عام 1999 لوضع معايير وأشكال رسمية لنهج التنقيب عن البيانات. كانت نتيجة عملهم هي CRISP-DM التي تمثل عملية استخراج البيانات القياسية في الصناعة، في عام 2001 كتب Jiawei Han & Micheline Kamber عن التنقيب عن البيانات باعتبارها جزءا رئيسا من اكتشاف المعرفة.

يعتبر التنقيب عن البيانات الضخمة المكون الأول من المكونين الأساسيين للبيانات الضخمة؛ أذ تتشكل البيانات الضخمة من التنقيب عن البيانات Big data mining موضوع هذه الورقة والمكون الآخر التحليلات التنبؤية Predictive Analytics، انظر (Dasgupta, January 2018, p. 22)

يرتبط التنقيب عن البيانات بكل ما هو متعلق باستخراج المعرفة المفيدة من كميات كبيرة من البيانات. إنه العثور على مجموعة صغيرة من الشذرات الثمينة (المعرفة) من قدر كبير من المواد الخام (البيانات) كما أشارت ادبيات البحث الى كم هائل من التعاريف لمفهوم تنقيب البيانات أشهرها تلك التي أشار لها موقع مؤسسة IGI Global يقدم 82 تعريفاً لمفهوم التنقيب في البيانات.

انطلاقا من هذه التعاريف يمكن تقديم بعض التعاريف المستنبطة:

• عملية البحث داخل كميات كبيرة من البيانات للكشف عن العلاقات التي لم يتم كشفها سابقا بين عناصر البيانات؛ والمعروفة أيضا باسم اكتشاف المعرفة في قواعد البيانات (KDD)

• هي النشاط الذي يقوم باستخراج المعلومات المتواجدة في كميات كبيرة من البيانات، بهدف البحث عن أنماط معرفية واكتشاف الحقائق الخفية الواردة في قواعد البيانات.

• هي عملية تحليل البيانات لتحديد العلاقات التي لم تكتشفها التحليلات السابقة من قبل. كما أنها تحليل البيانات لإقامة علاقات وتحديد أنماط.

• تحليل البيانات المتواجدة في قواعد البيانات باستخدام الأدوات التي تبحث عن الاتجاهات أو البيانات التي لا معنى لها، واستخراج معلومات ضمنية، لم تكن معروفة سابقا، ويمكن أن تكون مفيدة.

• عملية استخراج أنماط معرفية من مجموعات البيانات الكبيرة من خلال الجمع بين الأساليب من الإحصاءات والذكاء الاصطناعي مع إدارة قواعد البيانات.

• هو خطوة واحدة في اكتشاف المعرفة داخل قواعد البيانات (KDD) حيث تستخدم أساليب التحليل مثل: الشبكات العصبية Neural Networks أو الخوارزميات الجينية Genetic Algorithms أو شبكة القرارات Decision Trees والتقنية والنماذج الهجينة Hybrid Models لتحديد الأنماط والعلاقات في مجموعات البيانات.

• عملية اكتشاف المعرفة في قواعد البيانات عن طريق تحديد الأنماط والاتجاهات في البيانات التي تم جمعها باستخدام طرق مختلفة مثل : التصنيف Classification أو التحليل التسلسلي Sequential analysis أو العنقدة (التجميع) clustering أو قواعد الارتباط Association Rule

• عملية يتم من خلالها فرز كميات كبيرة من البيانات بهدف استخلاص المعلومات ذات الصلة. ويستخدم هذا المصطلح على نحو متزايد في العلوم لاستخراج المعلومات من مجموعات البيانات الهائلة الناتجة عن الطرق التجريبية والرصدية الحديثة، ولاسيما في سياق الحقول البيولوجية.

• عملية الاختيار، والاستكشاف، ووضع نماذج لكميات كبيرة من البيانات لاكتشاف الاطراد أو العلاقات المجهولة بهدف الحصول على نتائج واضحة ومفيدة لمالك قاعدة البيانات.

• فرع من علوم الكمبيوتر، يهتم باستخراج أنماط من مجموعات البيانات الكبيرة من خلال الجمع بين التحليل الإحصائي والذكاء الاصطناعي مع إدارة قواعد البيانات. ويعتبر استخراج البيانات كأداة ذات أهمية متزايدة من قبل رجال الأعمال الحديثة لتحويل البيانات إلى معلومات لإعطاء الأعمال ميزة إعلامية. ويستخدم حاليا في مجموعة واسعة من الممارسات التنميط، مثل التسويق، والمراقبة، والكشف عن الغش، واختيار شريك تجاري.

• فرع مهم في الصناعة والسوق، واسترجاع المعلومات الهامة من كمية هائلة من البيانات. ويستخدم عادة مع كمية هائلة من البيانات غير المتجانسة، حيث استخدام أجهزة الكمبيوتر أمر لا مفر منه

• عملية تقنية تستخدم خوارزميات لتحليل البيانات من وجهات نظر متعددة واستخراج أنماط ذات معنى والتي يمكن استخدامها للتنبؤ بالسلوك المستقبلي للمستخدمين. نظام تحليل سلة السوق الذي يستخدمه Amazon.com يوصي بمنتجات جديدة لعملائها على أساس مشترياتهم الماضية، وهو مثال معروف على نطاق واسع للكيفية التي يمكن أن تستخدم التنقيب عن البيانات في مجال التسويق.

• التنقيب عن البيانات هو حقل متعدد التخصصات، يستفيد من المجالات بما في ذلك تقنية قاعدة البيانات، الذكاء الاصطناعي، والتعلم الآلي، والشبكات العصبية، والإحصاءات والتعرف على الأنماط، والنظم القائمة على المعرفة، واكتساب المعرفة، واسترجاع المعلومات، والحوسبة عالية الأداء والصورة ومعالجة الإشارات، وتحليل البيانات المكانية و البيانات التصورية **(Data Visualization)**، والتي تعتمد بشكل كلي على الإدراك البصري.

وما يتم استنتاجه من التعاريف السابقة أنه، غالباً ما يتم الإشارة إلى التنقيب عن البيانات بصفة عامة إلى فكرة التحقيق بعمق في جبال من البيانات. هذا الاستخدام غير الرسمي لهذا المصطلح عادة ما يستخدم القليل من التقنيات للقيام بالتحقيق. في المقابل، فإن الاستخدام الرسمي لهذا المصطلح يشير تحديدا إلى استخدام التقنيات الحاسوبية للكشف عن أنماط في مجموعات البيانات الضخمة. تتنوع هذه التقنيات ما بين الأساليب الإحصائية إلى الذكاء الاصطناعي. ومجموعات التحقق واستخراج البيانات، وهي متنوعة ومتزايدة أكثر من أي وقت مضى. كما أنه هناك مصطلحات أخرى ذات الصلة التي تستخدم أيضا في الأوراق الأكاديمية والبحثية للتنقيب عن البيانات، مثل التنقيب عن المعرفة، واستخراج المعرفة، وأنماط تحليل البيانات، وعلم آثار البيانات وتحريف البيانات، ولكن في عالم الأعمال عموما مصطلح "التنقيب البيانات" هو أكثر شعبية،

فقد ظهر التنقيب عن البيانات في أواخر الثمانينيات وأثبت أنه أحد الحلول الفعالة لتحليل كميات هائلة من البيانات، وتحويلها من مجرد معلومات (بيانات) متراكمة وغير مفهومة إلى معلومات قيمة يمكن استغلالها واستخدامها لاحقًا هذا ولقد حظيت مرحلة التنقيب عن البيانات باهتمام كبير في مجتمع البحث على مدار العقد الماضي، بهدف تطوير خوارزميات قابلة للتطوير وقابلة للتكيف مع كميات متزايدة من البيانات بحثًا عن أنماط معرفية ذات مغزى حيث نمت حزم الخوارزميات والبرمجيات بشكل هائل على مدار العقد الماضي ، لدرجة أن التوسع جعل من الصعب على العاملين في هذا المجال مواكبة التقنيات المتاحة لحل مهمة معينة. كما تم استخدام شركات التنقيب؛ لزيادة الإيرادات وتحسين الإنفاق واستهداف عملاء جدد وتوفير خدمة عملاء أفضل والاستماع إلى ما يقوله الآخرون ويفعلونه من خلال الذكاء التنافسي (الأكاديمية، 2021) ويمكن تقسيم الأهداف من التنقيب على البيانات:

- من أجل تعليل بعض الظواهر المرئية. مثال: لماذا زادت نسبة المدخنين في الوطن العربي؟
- من أجل التثبت من نظرية ما. مثال: التثبت من النظرية التي تقول بأن الأسر الكبيرة تهتم بالضمان الصحي أكثر من الأسر الصغيرة عددا.
- من أجل تحليل البيانات للحصول على علاقات جديدة وغير متوقعة. مثال: كيف سيكون الانفاق العام إن كان ملازما لعمليات خداع واسعة من قبل البطاقات الائتمانية

2.2 مراحل التنقيب في البيانات:

لا يعد اكتشاف المعرفة في قاعدة البيانات (Knowledge Discovery in Database KDD) عملية سهلة، والتي يقول البعض إنها تنتهي بجمع البيانات وإدارتها كونه تتقاطع وتشترك في أمور عدة مع تحليل البيانات، بل هو خطوة وعملية صعبة ومثيرة للاهتمام (Che, 2013, April).

يعد التنقيب عن البيانات جزءًا من اكتشاف معرفة الاكتشاف مر بعدة مراحل وهي:

1.2.2 اكتشاف البيانات Data discovery: تتضمن مرحلة جمع البيانات اكتشاف وتحديد وتوصيف البيانات المتاحة.

2.2.2 تصفية البيانات وتنقيتها Data cleaning وتنقية البيانات يتم في هذه المرحلة إزالة البيانات المزعجة. التي ليس لها اهمية، وحذف جميع البيانات المتضاربة والغير المنسقة.

3.2.2 تكامل البيانات Data integration في هذه المرحلة، يتم جمع البيانات المتشابهة ويتم جمع البيانات ذات الصلة من مصادر متعددة ودمجها.

4.2.2 اختيار البيانات Data selection: في هذه المرحلة، يتم تحديد واسترجاع. البيانات المناسبة من مجموعة البيانات

5.2.2 تحويل البيانات Data transformationيتم تحويل البيانات إلى نماذج مخصصة مناسبة لإجراءات البحث والدراسة. ردود الفعل من خلال موضوع النجاح أو عملية التجميع.

6.2.2 التنقيب عن البيانات Data mining أي استخدام الأساليب الذكية المطبقة لاستخراج نماذج البيانات واستخراج النماذج المفيدة قدر الإمكان.

7.2.2 تقييم النمط Pattern evaluation، يتم تقيم النماذج المهمة جداً والتي تمثل قاعدة المعرفة من أجل استعمال بعض المقاييس المهمة.

8.2.2 تمثيل المعرفة وتقديمها Knowledge presentation هذه هي الخطوة الأخيرة لاكتشاف المعرفة في قواعد البيانات، وهي الخطوة التي يراها المستخدم. تستخدم هذه الخطوة الأساسية الطريقة المرئية لمساعدة المستخدم فهم وتفسير نتائج استخراج البيانات.

3.2: علاقة تنقيب البيانات والعلوم الأخرى

يُعتبر تنقيب البيانات ملتقى الجهود المبذولة من الباحثين في عدة مجالات من المعرفة، والذي من خلاله يتم تطوير وبناء تقنيات تتعامل مع البيانات وأشكالها المتعددة وأنواعها المختلفة بهدف مواجهة المشكلات في مجالات مختلفة كالهندسة، الأعمال، الصناعة، الطب والعلوم.

يجمع تنقيب البيانات بين عدة علوم كالإحصاء وتعليم الآلة وقواعد البيانات وتقنيات الإظهار المرئي، ويتجلى هذا الجمع في مراحل وخطوات تنقيب البيانات بدءاً من تجهيز وتجميع البيانات وحتى النتيجة النهائية والتي تختلف حسب أهداف وأغراض التنقيب. ويمكن تلخيص ما تم ذكره في الشكل الآتي:

الشكل رقم 2 علاقة تنقيب البيانات بالعلوم الأخرى

المصدر(الحميصي، 2017)

فبعض العمليات والطرق في تنقيب البيانات تأتي من علم الإحصاء، مثل نظرية naïve bays وmaximum entropy (مصطفى، 2017) وتقييم احتماليات التنبؤ، ويختلف تنقيب البيانات عن الإحصاء في أنه لا يتعامل فقط مع البيانات التي تحتوي أرقام فقط، فهناك بيانات بأشكال مختلفة، كما أن تنقيب البيانات يختلف في نوعية الطرق التي يتعامل فيها مع البيانات، حيث تميل هذه الطرق لمجال تعليم الآلة أكثر من الإحصاء الذي يتعامل مع الفرضيات وكم البيانات، كما أن تنقيب البيانات يستخدم بعض تقنيات تعليم الآلة مثل الشبكات العصبية وشجرة القرار، ويختلف هدف تنقيب البيانات عن هدف تعليم الآلة، فتعليم الآلة هدفه إعطاء الحواسيب القدرة على تنفيذ مهام يقوم بها البشر عبر تعليمها، بمعنى آخر: استبدال الدور البشري، ولكن تنقيب البيانات هدفه مساعدة الدور البشري ودعمه وليس استبداله، يتمثل الدور الرئيسي لقواعد البيانات في حفظ البيانات

والحصول عليها عند الحاجة، بينما دور تنقيب البيانات يتمثل في القدرة على قراءة هذه البيانات وتحليلها للمساعدة في اتخاذ القرار المناسب. تقنيات الإظهار المرئي يتم استخدامها كأداة في مرحلة تحضير البيانات أو مرحلة ما بعد تنقيب البيانات لإظهار النتائج.

3.مهام و تطبيقات تنقيب البيانات

1.3 مهام تنقيب البيانات

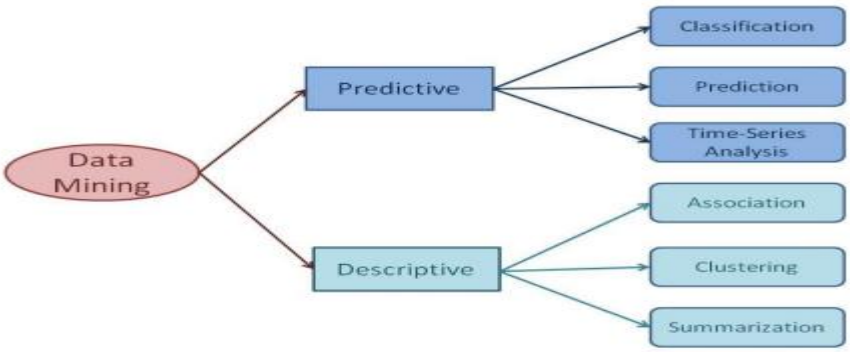
مهام تنقيب البيانات تُعَبِر عن طبيعة ونمط المعرفة التي نُنقب عنها ويُحدد ذلك طبيعة البيانات والخوارزميات المستخدمة في التنقيب، وتنقسم مهام تنقيب البيانات إلى نوعين: (الحميصي، 2017)

1.1.3 مهام تنبؤيه (Predictive) :

هدف هذه المهام هو القيام بعملية تنبؤ قيمة سمة من سمات البيانات استنادًا لقيم سمات أخرى. السمة التي يتم تنبؤ قيمتها تسمىTarget Class ، ويوجد نوعان لهذه المهام، الأول يُسمى التصنيف (Classification) والثاني يُسمى بالتنبؤ.(Prediction)

2.1.3 مهام وصفية :(Descriptive) هدف هذه المهام هو القيام بعملية وصف لخصائص البيانات إما عن طريق تجميع البيانات المتشابهة في خصائصها ضمن مجموعات ويسمى هذا النوع بتحليل العناقيد(Clustering) ، أو عن طريق إيجاد روابط وقواعد علاقات بين البيانات ويسمى هذا بتحليل الاقتران وينتج عنه قواعد تسمى .(Association Rules) يوجد نوع ثالث من المهام الوصفية ومهمته إيجاد البيانات الشاذة أو المشوهة ويسمى بتحليل القيم الشاذة (Outliers Analysis)

الشكل رقم 3 مهام تنقيب البيانات



المصدر (محمد، 2021)

2.3 تطبيقات تنقيب البيانات

يوجد العديد من الأمثلة على تطبيقات تنقيب البيانات، ففي الجانب التربوي مثلاً، يُستخدم تنقيب البيانات في توقع الأداء الدراسي للطلاب استناداً على بيانات أداء طلاب سابقين، وكذلك من الممكن التنبؤ بأداء المدرسين أو قدرتهم على إعطاء مساقات معينة، ويوجد العديد من الدراسات والأبحاث التي أجريت في هذا المجال، ويُطلق على تطبيقات تنقيب البيانات في المجال التربوي Educational Data Mining .

1.2.3 في مجال الأعمال والمال، يُستخدم تنقيب البيانات مثلاً في توقع قدرة العملاء على سداد قروضهم المالية، وهذا يتم بتطبيق خوارزميات تنقيب البيانات على سجلات تاريخية لعملاء سابقين، لينتج عنها نموذج أو مجموعة قواعد تحدد هل سيستطيع العميل سداد قرضه أم لا.

2.2.3 يوجد في مجال الطب أيضاً حضور واسع لتنقيب البيانات يتمثل في الحلول التي تقدمها كتوقع إصابة المريض بأمراض محددة وفقاً لبياناته وسجله الطبي، أو مدى تأثير الأدوية والعقاقير على المرضى بناءً على سجلات لمرضى سابقين. من أهم الأدوات المستخدمة في مجال الطب التنقيب عن البيانات وخاصة في مجال الاستكشاف ومعرفة الظروف الصحية السائدة، البحث في الأمراض وسلوك المريض في المجتمع، وكذلك المساهمة في معرفة تأثير الأدوية والعقاقير على المرضى حسب سجلات المرضى السابقة. كما يساعد التنقيب عن البيانات في التعرف على الأمراض المنتشرة في منطقة معينة، مما يساعد على اتخاذ الإجراءات والوعي اللازم للسيطرة على هذا المرض، ويعمل على تطوير وتحديث مجال الطب والأدوية وزيادة انتشارها. الكفاءة وقدرتها على المعالجة، فتحول البشر شيئاً فشيئاً من طب التعامل مع الأمراض إلى طب يسعى للتنبؤ بالأمراض ومنعها، وتقديم العلاج الملائم لكل شخص بمساعدة كم هائل من المعلومات يجمعها هاتفه الذكي، ما يفتح الباب لعصر جديد من الطب تلعب فيه البيانات الصحية الضخمة وتحليلاتها دوراً بارزاً. أصبح بإمكان المستشفيات الحكومية والمراكز الطبية والأطباء الاستفادة من البيانات الضخمة في دراسة سلوكيات المرضى عبر تحليل ملفاتهم الطبية والزيارات التي قاموا بها للعلاج والتقنيات القابلة للارتداء مما قد يساعدهم على تقديم خدمات طبية أفضل. يتيح تحليل البيانات الضخمة معالجة أوجه القصور في نظم تقديم الرعاية الصحية التي تتزايد تكاليفها بفعل النمو السكاني وارتفاع متوسط الأعمار. كما تستفيد العديد من المستشفيات في مختلف أنحاء العالم من البيانات الضخمة في تقليل وقت الانتظار في أقسام الطوارئ وتتبع حركة المريض، وزيادة كفاءة الإدارة الطبية. كما تستخدم البيانات الضخمة في صناعة الأدوية وتوزيعها وبيعها. تجمع شركات تصنيع الأدوية والتأمين الصحي البيانات من الدول في أفريقيا وآسيا مثلا لاستخدامها في التنبؤ بظهور أمراض معينة، وزيادة مبيعاتها في مناطق معينة، إذ تعتمد سياسات التسعير وتوزيع الأدوية على نتائج تحليل هذه البيانات (الأكاديمية، 2021).

3.2.3 في مكافحة الجريمة والاحتيال تُقدم تقنيات تنقيب البيانات حلولاً مهمة في تحليل بيانات المجرمين وسجلهم الإجرامي وتقوم بالتنبؤ مثلاً بإمكانية قيام أشخاص بجرائم معينة، والعمل على تحييدهم.

4.2.3 استخدام التعليم والتنمية البشرية لتطبيقات تنقيب البيانات إن لتنقيب البيانات دور محوري في كافة المجالات العلمية والعملية، وأصبح وجودها من ضرورات الأنظمة والتطبيقات. يتم استخدام التنقيب عن البيانات في الجانب التعليمي. من خلال التنبؤ بالأداء الأكاديمي للطلاب بناءً على بيانات من الطلاب السابقين. يتم أيضًا تحليل المناهج ونتائج الاختبارات من أجل استكشاف والتعرف على الاتجاهات العلمية للطلاب وتفضيلاتهم وتوافقهم. مع المحتوى والبرامج الأكاديمية، بالإضافة إلى تحليل النظام التعليمي بشكل عام، وإيجاد طرق لتطوير آلية العمل، ورفع مستوى أداء وخبرة المعلمين، والعمل على حلها. المشاكل التي تواجه التنمية بشكل عام في المجتمع، وتسمى هذه التطبيقات التنقيب عن البيانات التعليمية التي تهدف إلى دعم اتخاذ القرارات المستنيرة والجهود المبذولة لتحسين الفعالية التعليمية انظر (Fischer, et al., 2020)

5.2.3 استخدام الهندسة الوراثية لتطبيقات تنقيب البيانات كذلك يساهم التنقيب عن البيانات بشكل أساسي في علم الهندسة الوراثية والمعلوماتية الحيوية نظرًا لكمية البيانات الكبيرة التي تحتويها والتي تكاد لا تنتهي. لقد عملت وسائل الاستكشاف بفاعلية في دراسة وتحليل السلاسل الجينية لاكتشاف الأنماط وحل الألغاز. الهندسة الوراثية للإنسان والتي ساعدتنا على فهم العديد من الأمراض وإيجاد العديد من الأدوية.

6.2.3 استخدام الفضاء وعلم الفلك لتطبيقات تنقيب البيانات أدوات الاستكشاف هي أفضل طريقة لتحويل الكمية الكبيرة من البيانات الجديدة الناتجة عن استكشاف أبحاث الفضاء وعلم الفلك والطبيعة وكذلك التطور التكنولوجي الهائل لتقنيات التصوير الرقمي والعدسات والتلسكوبات، والاستفادة منها في فهم الطبيعة، حيث يزيد ذلك من قدرة الإنسان على التعامل مع التغيرات استكشاف الأسرار الطبيعية والمناخية الكامنة وراءها طبقات الغلاف الجوي المختلفة .

7.2.3 التطبيقات العامة في مجالات المجتمع تعد المزايا المختلفة لتقنيات التحليل المتقدمة واستخراج البيانات في مختلف المجالات الحيوية من بين أهم الطرق لتمكين تحليل أكبر قدر من البيانات المتعلقة بمجالات متعددة معًا بطريقة متكاملة، من أجل استكشاف العلاقات والترابط. ومدى التأثير المتبادل بين هذه المجالات بحيث تتوصل إلى استنتاجات تساهم في صنع القرار بشأن المواقف السياسية واستراتيجيات التنمية والتنمية المهمة)أبرار , 2021) حيث يمكن تطبيقات التنقيب في البيانات الاجتماعية، واستخراج البيانات بالطرق الآتية:

-1كتابة تقرير موجز عن مجموعة محددة سكان التنميط : هذا هو من خلال تطوير وبناء تقارير موجزة عن مجموعة من العملاء البارزين وعلى بطاقات الائتمان.

-2 أيضا تحليل الاتجاه التجاري : البحث والعثور على أسواق ذات خصائص نمو قوية أو ضعيفة.

-3 كذلك التسويق لغرض معين. التسويق المستهدف: البحث عن العملاء لغرض معين، مثل تقديم خصومات لهم.

-4 استخدام التحليل : ابحث عن طريقة محددة لاستخدام السلع والخدمات المختلفة.

-5 فعالية الحملة : اكتشف أكثر الحملات فاعلية وتأثيراً من خلال مقارنة الاستراتيجيات المختلفة لهذه الحملات (الأكاديمية، 2021).

3.3 تحديات تنقيب البيانات

نستطيع القول إن مجال تنقيب البيانات هو مجال بحثي حي ملئ بالمساهمات العلمية التي تُقدم حلولاً للتحديات والمشاكل في هذا المجال، وكما في أي مجال علمي وبحثي، يوجد تحديات بارزة تواجه تنقيب البيانات تتمثل في:

1.3.3 قابلية التوسع :(Scalability) لابد أن تتوفر في تقنيات تنقيب البيانات القدرة على التعامل مع كم كبير وضخم من البيانات.

2.3.3 **الخصائص الكثيرة** :(High Dimensionality) العديد من التطبيقات ينتج عنها كم كبير من الخصائص قد يصل للآلاف، وهذا يشكل تحدياً لتقنيات التنقيب.

3.3.3 البيانات الغير متجانسة والمعقدة :(Heterogeneous data) في السنوات الأخيرة بدأت أشكال جديدة للبيانات في الظهور تحتوي على رسومات، بيانات غير نصية وأنواع جديدة من البيانات، وينبغي على تقنيات تنقيب البيانات أن تتعامل مع هكذا أنواع من البيانات، ولمزيد من التفصيل حول تنوع وعدم تجانس انظر(Wu, 2013)

4.3.3 جودة البيانات (Data Quality) : البيانات الواقعية والحقيقية دائماً ما تحتوي على قيم فارغة وقيم غير منطقية أو شاذة. لمعالجة جودة البيانات، لابد من تطوير خوارزميات تنقيب تأخذ بعين الاعتبار مشكلة الجودة.

انتشار البيانات في أماكن متعددة :(Data Distribution) كلما زاد حجم البيانات، كلما كانت منتشرة في أماكن مختلفة، لذلك، يجب تطوير تقنيات تنقيب تتعامل مع بيانات منتشرة وغير موجودة في مكان واحد.

5.3.3 الحفاظ على خصوصية البيانات (Privacy Preservation): تعتني خصوصية البيانات بعدم نشر المعلومات الخاصة في البيانات، بينما هدف تنقيب البيانات هو استخراج المعرفة، لذا يوجد اهتمام متزايد في تطوير خوارزميات للتنقيب تحافظ على خصوصية البيانات.

4. الخاتمة:

نلخص في نهاية هذه الورقة البحثية مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات ومخازن البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات واستخراج المعلومات والمعارف منها، من هنا ظهر ما يسمى بالتنقيب في البيانات كتقنية تهدف إلى استخراج المعرفة من كميات هائلة من البيانات. وهي تقنية حديثة فرضت نفسها بقوة في عصر المعلوماتية، واستخدامها يوفر للشركات والمنظمات في جميع المجالات القدرة على استكشاف والتركيز على أهم المعلومات في قواعد البيانات، من أهم نتائج هذه الدراسة بيان المراحل المختلفة للتنقيب عن البيانات الضخمة وتحديد أبرز التحديات؛ حيث يبقى التنقيب عن البيانات الضخمة معقدا في ظل البيانات غير المتجانسة مع صعوبة التوفيق بين ثنائية جودة البيانات والحفاظ على الخصوصية.

5. قائمة المراجع

المراجع العربية

أبرار, م .(16 01, 2021) .تاريخ التنقيب في البيانات Récupéré sur Big Data Analytics What it is and why it matters: https://bit.ly/3Pbhjf8 .

الأكاديمية, م. ا .(2021) .مفهوم التنقيب في البيانات وتقنياته المختلفة Récupéré sur بيانات : https://www.baianat.com/ar/articles/mine-gold-in-your-data

الحميصي, ا .(08 10, 2017) .نمذجيات https://bit.ly/3Fv2mQS: Récupéré sur https://www.nmthgiat.com/ .

حايك, ه Récupéré sur (27 03, 2014) .التنقيب في البيانات Data mining واستخراج المعرفة https://bit.ly/3N4Ibvs :

سعد مأمون, ع & ,.أحمد عثمان, ع .(2015) .أهمية ودور الرقم الوطني في تقليص البيانات الضخمة .المجلة العربية الدولية للمعلوماتي , 35-40.

طيار أحسن، شلابي عمار .(s.d.) .التنقيب في البيانات و اتخاذ القرارات.

قرزيز, أ .(2021) .تنقيب البيانات الضخمة يف املكتبات الرقمية: دراسة يف املفاهيم . مجاة ببليوفيليا لدراسات المكتبات والمعلومات.215-194 ,

محمد, ا .(16 01, 2021) .تاريخ التنقيب على البيانات Récupéré sur .تاريخ التنقيب على البيانات : https://bit.ly/3Fv2mQS

مصطفى, ف. ع .(2017) .التحليل المتقدم للتنقيب عن البيانات .لبنان: دار الفكر العربي.

المراجع الأجنبية

Che, D., Safran, M., & Peng, Z. (2013, April). From big data to big data mining: challenges, issues, and opportunities. In International conference on database systems for advanced applications (pp. 1-15). Springer, Berlin, Heidelberg.

Dasgupta, N. (January 2018). Practical Big Data Analytics. Livery Place 35 Livery Street, Birmingham: Packt Publishing Ltd.

Fischer, C., Pardos, Z. A., Ryan Shaun , B., Joseph Jay , W., Padhraic , S., Renzhe , Y., . . . Mark , W. (2020). Mining Big Data in Education: Affordances and. Dans F. e. al, Big Data in Education. National Science Foundation: Review of Research in Education, 44.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2013). Data mining with big data. IEEE transactions on knowledge and data engineering, 26(1), 97-107.