جامعة بنها
كلية الآداب

مجلة كلية الآداب

مجلة دورية علمية محكمة

التنقيب في البيانات الماهية والأهمية: مدخل نظري

إعداد /
هند محمد يوسف عقدة

المجلد ٦١ ابريل ٢٠٢٤
https://jfab.journals.ekb.eg/

الملخص

تتناول الدراسة إلقاء الضوء على أهمية التنقيب في البيانات والذي نشأ نتيجة الحاجة إلى وسيلة يمكن من خلالها حفظ الكمية الهائلة من البيانات الموجودة وتنظيمها ومعالجتها، ولذلك تناولت الدراسة مهام ووظائف التنقيب في البيانات، والتقنيات المبتكرة التي ظهرت لكي تعمل على ضبط وتحسين مفاهيم التنقيب في البيانات في محاولة منح المؤسسات رؤية أكثر شمولا حول بياناتها، بالإضافة إلى الأدوات الخاصة بها سواء مفتوحة المصدر أو تجارية، كما تناولت الباحثة أهمية التنقيب في البيانات في المؤسسات ومراكز المعلومات مع توضيح تطبيقات التنقيب في البيانات في المكتبات ومراكز المعلومات .

وتوصلت الدراسة إلى مجموعة من النتائج ومن أهمها ما يلي:

- تبين في الآونة الأخيرة في مجال المكتبات والمعلومات، وجود زيادة في عدد الوظائف والتي تهتم بالبيانات بشكل عام، حيث تنوعت المسميات الوظيفية لتشمل:(مختص البيانات، ومكتبي البيانات، ومنسق البيانات، وأرشيفي البيانات، وعالم البيانات)، كما لوحظ زيادة في أعداد التسميات الوظيفية ذات الصلة بالبيانات على سبيل المثال:(مكتبي خدمات البيانات البحثية، ومكتبي إدارة البيانات البحثية، ومكتبي الوعي البياناتي، ومكتبي البيانات البحثية، ومكتبي معالجة البيانات) وغيرها من تسميات ذات صلة.
- يحتاج اختصاصيو البيانات في الوقت الحالي إلى مهارات تقنية لاستخراج البيانات وتحليلها، يحتاج أيضا إلى الجمع بين المهارات والمعرفة المطلوبة لإدارة البيانات، ومعالجة البيانات، والكفاءات لدعم إدارة البيانات.

- لايحتاج اختصاصي البيانات أن يصبح مبرمجاً ، ولكن يجب أن يهتم بمعرفة اللغات ومنطق البرمجة لأجهزة الكمبيوتر. وبالتالي ، سوف يكون على دراية بتشغيل مجموعة من البرامج العلمية المستخدمة لتحويل البيانات.

الكلمات المفتاحية

التنقيب في البيانات Data Mining، تعلم الآلة Machine Learning، اكتشاف المعرفة في قواعد البيانات Knowledge Discovery in databases، تطبيقات التنقيب في البيانات في المكتبات، أدوات التنقيب في البيانات، علم البيانات Data Science

المقدمة

أدى الانتشار الواسع لتكنولوجيا المعلومات وسهولة توفرها إلى زيادة في حجم البيانات التي لم نشهدها من قبل في التاريخ ، مما جعل قضية البيانات الضخمة على الإنترنت موضع جدل، من حيث عن جدوى وجودها بهذه الطريقة العشوائية، وعندما نتحدث عن البيانات الضخمة ، فإننا نتحدث عن كميات لا يمكن تصورها من البيانات من أنواع ومصادر متعددة بحجم مئات تيرابايت أو حتى بيتابايت.[١]

بناءً على ما سبق ، نشأت حاجة كبيرة إلى وسيلة يمكن من خلالها حفظ الكمية الهائلة من البيانات الموجودة وترتيبها وتنظيمها ومعالجتها ، مما أدى إلى ظهور ما يعرف اليوم باسم التنقيب في البيانات[٢]. تستند فكرة التنقيب في البيانات إلى الاستنتاجات المعرفية لكميات كبيرة من المعلومات بناءً على خوارزميات رياضية

[1] Stancin, Igor and Alan Jović. "An overview and comparison of free Python libraries for data mining and big data analysis." *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2019): 978.

[2] C ,Chen, Xing, Z., & Liu, Y.. What's spain's paris? mining analogical libraries from q&a discussions. Empirical Software Engineering, 24(3), (2019), 1160.

مستمدة من الإحصاء والرياضيات والذكاء الاصطناعي والعديد من العلوم الذكية وغير التقليدية الأخرى[١].

أهمية الدراسة

١. التعريف بمفهوم التنقيب في البيانات.
٢. توضيح أدوات وتقنيات التنقيب في البيانات.
٣. استنباط أهمية تقينات التنقيب في البيانات في تخصص المكتبات والمعلومات.
٤. اكتشاف تطبيقات التنقيب في البيانات في المجالات المختلفة بصفة عامة ومجال المكتبات بصفة خاصة.

تساؤلات الدراسة

١. ما المقصود بتقنية التنقيب في البيانات؟
٢. ما أدوات التنقيب في البيانات؟
٣. ما مدى حاجة تخصص المكتبات والمعلومات إلى تقنيات التنقيب في البيانات؟
٤. ما واقع تطبيق تقنيات التنقيب في البيانات في المجالات المختلفة؟

أهداف الدراسة

١. دراسة ماهية التنقيب في البيانات وأهميته.
٢. التعرف على أدوات التنقيب في البيانات.
٣. استكشاف حاجة تخصص المكتبات والمعلومات إلى التنقيب في البيانات
٤. الوصول إلى تطبيقات التنقيب في البيانات في المجالات المختلفة.

[1]) S ,Li, et. Problems and changes in digital libraries in the age of big data from the perspective of user services . The Journal of Academic Librarianship, 45(1), (2019). 22.

منهج الدرسة وأدواتها

تعد هذه الدراسة من الدراسات الوثائقية التي تقوم على جمع المعلومات من المصادر والمراجع الموثوقة لأخذ المعلومات الدقيقة، والمتعلقة بالبحث منها، والتي تساهم في الوصول إلى النتائج التي يسعى البحث لتحقيقها.

أداة الدراسة:

الإنتاج الفكري وأدب الموضوع، حيث سيتم الرجوع إلى الدراسات المتخصصة وغير المتخصصة التي تتعلق بعلم البيانات Data Scienc والتنقيب في البيانات Data Miningبشكل عام وفي تخصص المكتبات والمعلومات بشكل خاص، بشقيه الإنتاج الفكري باللغة العربية، واللغة الإنجليزية من خلال الاستناد إلى الكتب والأطروحات العلمية وأعمال المؤتمرات العلمية بالإضافة إلى مقالات الدرويات المتخصصة.

الداسات السابقة

الدرسات العربية

أسماء قرزيز. تنقيب البيانات الضخمة في المكتبات الرقمية: دراسة في المفاهيم وآليات التطبيق. مجلة ببليوفليا لدراسات المكتبات والمعلومات . مج٠٣،ع ١٠. ٢٠٢١.

تسعى هذه الدراسة إلى تكوين خلفية نظرية لمفهومي البيانات الضخمة وتنقيب البيانات، وتحليل وتوضيح مجالات استخدام أدوات التنقيب عن البيانات الضخمة في المكتبات الرقمية، التحسيس بإلزامية تفعيل مشاريع واقعية لاستغلال تطبيقات وفوائد تنقيب البيانات الضخمة في المكتبات الرقمية، بالإضافة إلى عرض تجربتين لمكتبتين رقميتين طبقتا تقنية البيانات الضخمة.

فاطمة يحيى زكريا، القضايا الأخلاقية والقانونية لإستخدام بيانات المستفيدين للتنقيب عن البيانات في المكتبات المصرية: دراسة تحليلية، البوابة العربية للمكتبات والمعلومات، ع٥، سبتمبر ٢٠٢٠.

تهدف الدراسة التعرف على القيم الأخلاقية التي نصت عليها المواثيق الدولية بشأن حماية بيانات المستفيدين في المكتبات، وإثارة الوعي بالعوائق القانونية في حالة استخدام بيانات المستفيدين الشخصية للتنقيب عن البيانات في المكتبات، بالإضافة إلى التعرف على أبرز الأساليب المستخدمة لتخفيف آثار التنقيب عند استخدام بيانات المستفيدين.

الدراسات الأجنبية

Mohammad Ahmad Niqresh. The Influence of Data Mining in Increasing Benefits of Libraries in Jordanian Governmental Universities" Library Philosophy and Practice (e-journal). (2021).

هدفت هذه الدراسة إلى دراسة أثر التنقيب في البيانات على زيادة الاستفادة من المكتبة الجامعية في الجامعات الحكومية الأردنية. واعتمد الباحث تقنيات تنقيب البيانات والتي شملت (قواعد الربط ، التصنيف، التجميع، التنبؤ، الأنماط المتسلسلة، وأشجار القرار). من خلال استخدام النهج الكمي واستخدام الاستبيان كأداة للدراسة، حيث ظهر تأثير التنقيب عن البيانات في تحسين وتنظيم وتدفق وتراكم البيانات المكتبية وتحسين خدمات المكتبة للمستخدمين. ومن بين تقنيات التنقيب عن البيانات ظهر أن (النمط المتسلسل وأشجار القرار وتقنيات التنبؤ) هي أكثر التقنيات تأثيراً على الخدمات المكتبية يليها أمناء المكتبات في تطوير الخدمات المكتبية.

Nasim Ansari, Hossein Vakilimofrad, Muharram Mansoorizadeh. Using data mining techniques to predict user's behavior and create recommender systems in the

libraries and information centers. Global Knowledge, Memory and Communication .Vol. 70 No. 6/7, (2021).

تهدف هذه الدراسة إلى تحليل سلوك المستفيد والتنبؤ به وإنشاء أنظمة التوصية في المكتبات ومراكز المعلومات باستخدام تقنيات التنقيب عن البيانات. وذلك من خلال مسح وتحليل البيانات الخاصة بالمكتبة حيث تم جمع البيانات المطلوبة لهذه الدراسة من معاملات المستفيدين بالمكتبات ومراكز المعلومات في جامعة همدان للعلوم الطبية. وباستخدام تقنيات استخراج البيانات، تم التحقيق في الأنماط الحالية، وتم تحليل معاملات قروض المستخدمين.

Wichai Puarungroj, Pathapong Pongpatrakant, Narong Boonsirisumpun, and Suchada Phromkhot, Investigating Factors Affecting Library Visits by University Students Using Data Mining. WKW School of Communication & Information & NTU Libraries, Nanyang Technological University. volume 28, issue 1, (2018).

تهدف هذه الدراسة إلى استخدام خوارزميتين لاستخراج البيانات وهما Naïve Bayes و شجرة القرار وذلك لتحليل البيانات. وكان الهدف من نتائج التنقيب في البيانات هو استخدامها في تشجيع الطلاب الجامعيين على زيارة المكتبة فعليًا. وأسفرت النتائج إلى تحديد سمات الطالب المفيدة للتنبؤ بزيارة المكتبة. ويمكن استخدام نتائج التنقيب في البيانات لزيادة معدل استخدام المكتبة من خلال تنظيم الأنشطة التي تستهدف تلك السمات.

التنقيب في البيانات: التعريفات والمفاهيم

على الرغم من أهمية التنقيب في البيانات المعترف بها، لا يزال هناك اختلافات في الآراء بشأن تعريفها وتطورت تلك التعريفات وفقا لما طرأ علي تقنية التنقيب في البيانات من تكنولوجيا، هناك مصطلحات أخرى ذات الصلة التي تستخدم أيضا في الأوراق الأكاديمية والبحثية للتنقيب في البيانات، مثل التنقيب عن المعرفة، واستخراج المعرفة، وأنماط تحليل البيانات، وعلم آثار البيانات وتجريف البيانات، ولكن في عالم الأعمال عموما مصطلح "التنقيب البيانات" هو أكثر شعبية.

- عملية لتحليل البيانات من وجهات نظر مختلفة وتلخيصها في معلومات مفيدة (معلومات يمكن استخدامها لزيادة الإيرادات أو خفض التكاليف أوكليهما).[١]
- دراسة جمع البيانات وتنظيفها ومعالجتها وتحليلها واكتساب رؤى مفيدة من البيانات.[٢]
- إنشاء علاقات (Relationships) بين البيانات أو فهم الأنماط (Patterns) الموجودة مسبقا.[٣]
- مجموعة من الأدوات التي تسهل استكشاف وتحليل البيانات الموجودة في "مستودع البيانات" أو " قواعد البيانات" .[٤]

[1]) Deepashri, K.S. and Kamath, A. ,"Survey on techniques of data mining and its applications", International Journal of Emerging Research in Management and Technology, Vol. 6 No. 2, (2017), pp. 198-201.

[2]) Bano, Shohar. And Gul, Sumeer, Exploring data mining: facets and emerging trends. Digital Library Perspectives
Vol. 37 No. 4, (2021). p. 431

[3]) Amin, M. S., Chiam, Y. K., & Varathan, K. D.. Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, (2019) 82-93.

[4]) Wang, S., Cao, J., & Yu, P. Deep learning for spatio-temporal data mining: A survey. IEEE Transactions on Knowledge and Data Engineering. (2020).

التنقيب في البيانات: مهامه ووظائفه

تعبر مهام تنقيب البيانات عن نمط وطبيعة المعرفة التي يتم التنقيب عنها، ويحدد ذلك طبيعة البيانات والخوارزميات المستخدمة في التنقيب، وتنقسم مهام تنقيب البيانات إلى نوعين كالتالي:

- **مهام تنبؤية (Predictive):** تستخدم المهمة التنبؤية متغيرات أو قيماً محددة في مجموعة البيانات للتنبؤ بقيم غير معروفة أو مستقبلية لمتغيرات أخرى ذات أهمية. [١] السمة التي يتم تنبؤ قيمتها تسمى Target Class، ويوجد نوعان لهذه المهام، الأول يُسمى التصنيف (Classification) والثاني يُسمى بالتنبؤ (Prediction).
- **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). هناك نوع ثالث من المهام الوصفية ومهمته إيجاد البيانات الشاذة أو المشوهة ويسمى بتحليل القيم الشاذة (Outliers Analysis).[٢]

تقنيات التنقيب في البيانات

هناك العديد من التقنيات المبتكرة التي ظهرت والتي تعمل على ضبط وتحسين مفاهيم التنقيب في البيانات في محاولة لمنح المؤسسات رؤية أكثر شمولاً حول بياناتها الخاصة مع اتجاهات مستقبلية جيدة، واستخدامها من حل العديد من المشكلات التي تواجه المؤسسات.

---

1) Ch, Tiju.. Astudy and survey of big data using data mining techniques.international journal of engineering sciences & research technology. 6(10): (October, 2017). P170.

2) Cherian, Tiju. Ibid, p171.

١) التصنيف Classification

التصنيف أسلوب تقليدي للتنقيب في البيانات يعتمد على التعلم الآلي بشكل أساسي، يتم استخدام التصنيف لتصنيف كل عنصر في مجموعة من البيانات إلى مجموعة محددة مسبقًا من الفئات أو المجموعات. تستخدم طريقة التصنيف التقنيات الرياضية مثل أشجار القرار(decision trees)، والبرمجة الخطية (linear programming)والشبكة العصبية (neural network) والإحصاءات(statistics). في التصنيف، يتم تطوير البرنامج الذي يمكنه تعلم كيفية تصنيف عناصر البيانات إلى مجموعات.[١]

وهي أكثر تقنيات التنقيب في البيانات شيوعًا، والتي تستخدم مجموعة من الأمثلة المصنفة مسبقًا لتطوير نموذج يمكنه تصنيف عدد السجلات بشكل عام. تُستخدم تقنية التصنيف في تقسيم العملاء ونمذجة الأعمال وتحليل الائتمان والعديد من التطبيقات الأخرى. على سبيل المثال ، تصنيف البلدان على أساس عدد السكان، أو تصنيف الدراجات على أساس الأميال.[٢]

٢) التجميع Clustering

إن طريقة التحليل العنقودي مشابه تقريبًا لمبدأ طريقة تحليل التصنيف. حيث يتم تقديم النتيجة النهائية في شكل عينات مجمعة. يحتاج التجميع أولاً إلى تجميع عدة عينات غير مصنفة وفقًا لأهم خصائصها، أي عينات من نفس الفئة لها نفس خصائص العينة تقريبًا، مع وجود اختلاف بين هذه الفئات.[٣]

إن تقنية العنقدة أو التجميع تختلف عن تقنيات التصنيف والتنبؤ. فهي عملية تبادلية مع طرق التصنيف، حيث يقوم التجميع بتحليل البيانات فقط دون مراعاة تسميات

1) Se, Sh. Applications of data mining in library & information centres: an overview. International Journal of Current Research Vol. 9, Issue, 01, (January2017) ,p45247.

2) P.Boopathi, L.Ramesh. A Study on Techniques of Data Mining and its Applications. International Journal of Creative Research Thoughts (IJCRT. Volume 6, Issue 1.(March 2018). P1237

3) Ibid.

الفئات الثابتة. قبل التجميع، لا يعرف عدد ونوع المجموعات المراد تقسيمها، ولا يعرف عناصر البيانات الذي يجب تحديد المجموعة وفقا لها. يتم تقسيم البيانات إلى مجموعات مختلفة، ويتم القيام بفروق بين المجموعات أكبر ما يمكن، والاختلافات داخل المجموعات صغيرة بقدر الإمكان .

٣) نماذج أو قواعد الربط Association

تشير قواعد الربط إلى انتظام معين بين قيم متغيرين أو أكثر. ظهور نمط واحد يعني ظهور نمط آخر. هو نوع مهم من الاكتشاف في قاعدة البيانات. فالغرض من المعرفة هو استكشاف العلاقات المتبادلة الغير مرئية في البيانات، بما في ذلك ارتباط التحقق، وترابط السلاسل الزمنية، والارتباط السببي. عادة ما تحتاج قواعد الربط إلى معرفة أن الدعم والثقة أكبر من أو يساوي الحد الأدنى من الدعم والثقة المحدد من قبل المستخدم.[١] الرابطة هي واحدة من أشهر تقنيات التنقيب في البيانات، يتم اكتشاف نمط بناءً على العلاقة بين العناصر في نفس المعاملة. هذا هو السبب في أن تقنية الارتباط تُعرف أيضًا باسم تقنية العلاقة. حيث يتم استخدام تقنية الارتباط في تحليل السوق لتحديد مجموعة من المنتجات التي يشتريها العملاء بشكل متكرر معًا، كما يتم تطبيقه لتوجيه منتجات تحديد المواقع داخل المتاجر بطريقة لزيادة المبيعات.[٢]

٤) التنبؤ Prediction

التنبؤ هو أحد تقنيات التنقيب في البيانات التي تكتشف العلاقة بين المتغيرات التابعة والمتغيرات المستقلة. على سبيل المثال، يمكن استخدام تقنية تحليل التنبؤ في البيع للتنبؤ بالربح في المستقبل إذا اعتبر البيع متغيرًا مستقلاً، يمكن أن يكون الربح متغيرًا

1) Ibid.

2) Se, Shantashree. Applications of data mining in library & information centres: an overview. International Journal of Current Research Vol. 9, Issue, 01, January,(2017),p45247

تابعًا. بعد ذلك، استنادًا إلى بيانات البيع والربح، يمكن رسم منحنى انحدار ملائم يستخدم لتنبؤ الربح.[١] من تطبيقات التنبؤ؛ الموافقة على الائتمان، التسويق، التشخيص الطبي، تحليل فعالية العلاج.[٢]

٥) الإنحدار Regression

الجانب الآخر للتنقيب في البيانات التنبؤية هو الانحدار، وهي تقنية خاضعة للإشراف للتنبؤ بهدف رقمي. في تقنية الانحدار يقوم بتقييم القيمة المستهدفة من حيث وظيفة تنبئ كل عنصر من عناصر البيانات. يتم بعد ذلك صياغة العلاقة بين القيمة المستهدفة والمتنبئين في نموذج يمكن تطبيقه على مجموعات بيانات مختلفة ذات قيم هدف غير معروفة ومنها ما هو انحدار خطي وانحدار لوجستي.[٣] يحاول العثور على دالة تقوم بنمذجة البيانات بأقل خطأ، لتقدير العلاقات بين البيانات أو مجموعات البيانات.[٤] حيث يقوم على تحديد وتحليل العلاقة بين المتغيرات. يتم استخدامه لتحديد احتمالية متغير معين، بالنظر إلى وجود متغيرات أخرى.[٥]

1) Sen, Shantashree. Applications of data mining in library & information centres: an overview. International Journal of Current Research Vol. 9, Issue, 01,(January 2017),p45247

2) P.Boopathi, L.Ramesh. A Study on Techniques of Data Mining and its Applications. International Journal of Creative Research Thoughts (IJCRT. Volume 6, Issue 1,(March 2018). P1238

3) ch, Tiju.. Astudy and survey of big data using data mining techniques. International journal of engineering sciences & research technology. 6(10): (October, 2017). p170.

4) G. Sangeetha Lakshmi, M. Jayashree. Comparative Analysis of Various Tools for Data Mining and Big Data Mining. International Research Journal of Engineering and Technology (IRJET). Volume 06 Issue 04 . (Apr 2019). P704.

5) R. Shankar, S. Duraisamy. Analysis of Data Mining Tasks, Techniques, Tools, Applications And Trends. IOSR Journal of Computer Engineering (IOSR-JCE). Volume 20, Issue 5, Ver. II (Sep - Oct 2018), p13.

٦) شجرة القرار Decision trees

تعد شجرة القرار واحدة من أكثر تقنيات التنقيب في البيانات شيوعا، لأن النموذج الخاص بها سهل الفهم للمستخدمين، في تقنية شجرة القرار يكون جذر شجرة القرار سؤالا أو شرطا بسيطا له إجابات متعددة، ثم تؤدي كل إجابة إلى مجموعة من الأسئلة أو الشروط التي تساعدنا في تحديد البيانات حتى نتمكن من اتخاذ القرار النهائي بناءا عليها.[١]

أهمية التنقيب في البيانات

يجب النظر إلى التنقيب في البيانات على أنه نتيجة للتطور الطبيعي لتكنولوجيا المعلومات. حيث شوهد مسار تطوري في قاعدة البيانات، في جمع البيانات وإنشاء قاعدة البيانات، وإدارة البيانات (بما في ذلك تخزين البيانات واسترجاعها، ومعالجة قاعدة البيانات)، وتحليل البيانات وفهمها (بما في ذلك تخزين البيانات واستخراج البيانات).[٢] ولذلك يجب توضيح أهمية الحاجه إلى التنقيب في البيانات (DM) في النقاط التالية:

١. باستخدام هذه التقنية، يتم تحليل البيانات ثم تحويل تلك البيانات إلى معلومات مفيدة. وهذا يساعد المؤسسات على اتخاذ أدق وأفضل القرارات.

٢. بمساعدة التنقيب في البيانات، يمكن تحليل سلوكيات العملاء ووجهات نظرهم. وهذا يؤدي إلى نجاح كبير وأعمال تعتمد على البيانات.

٣. لا تساعد عملية التنقيب في الكشف على العلاقات المفيدة الموجودة بين البيانات فحسب، بل تساهم أيضا في بناء ارتباطات وعلاقات جديدة.

[1]) Sen, Shantashree. Applications of data mining in library & information centres: an overview. International Journal of Current Research Vol. 9, Issue, 01,(January, 2017),p45247

[2]) Ka, Ba. Uma, Mu. An Overview of General Data Mining Tools. International Research Journal of Engineering and Technology (IRJET). 04 ,09 . (Sep -2017). p930.

٤. يساعد في الكشف عن مخاطر الائتمان والاحتيال.

٥. فهم الأحداث السابقة والتنبؤ الصحيح للأحداث المستقبلية وعلى ذلك العمل في إطار ما سيحدث.

أدوات التنقيب في البيانات

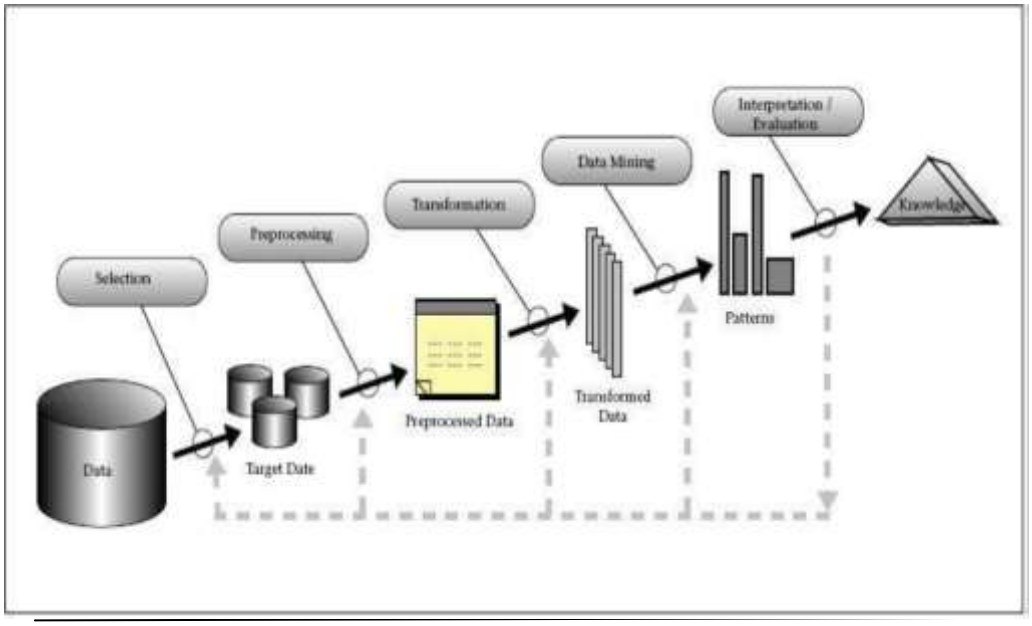
في الوقت الحاضر، يجعل التطور السريع للبيانات عملية اتخاذ القرار تكون معقدة وصعبة. قد يصبح التنقيب في البيانات هو الحل لأن مهمته اكتشاف المعرفة. ولذلك هناك الكثير من أدوات التنقيب في البيانات التي يمكن استخدامها لاتخاذ القرار كالتالي:

م	الأدوات مفتوحة المصدر Open Source	الأدوات التجارية COMMERCIAL
١	Weka	Rapid Miner
٢	Knime	Sisense
٣	Keel	Board
٤	Orange	
٥	Tanagra	
٦	R	
٧	Scikit-Learn	
٨	Spark	

خطوات عملية التنقيب في البيانات

يعرف التنقيب في البيانات (DM) أيضا باسم اكتشاف المعرفة في قاعدة البيانات Knowledge Discovery in Databases (KDD) ويتم أحيانا استخدام المصطلح بالتبادل ويشير إلى البحث عن المعرفة أو التنقيب عنها من كميات هائلة من البيانات، تستخدم تقنيات التنقيب في البيانات للعمل على كميات كبيرة من البيانات لاكتشاف الأنماط والعلاقات المخفية المفيدة في اتخاذ القرار. لذلك فإن العلاقة بين التنقيب في البيانات (DM) واكتشاف المعرفة(KDD) علاقة اشتمال من الأولى للثانية، حيث أن كينونة (KDD) تتضح في أنها عملية تهدف إلى اكتشاف المعرفة من واقع البيانات، في حين يشير (DM) إلى كونه خطوة أو مرحلة معينة في اطار عملية (KDD) ولذلك يوضح الشكل رقم (١) والجدول رقم (١) المراحل المختلفة لاكتشاف المعرفة في قواعد البيانات.

شكل رقم (٧) يوضح مراحل اكتشاف المعرفة في قواعد البيانات [١]



١/٩/١ مراحل عملية اكتشاف المعرفة (KDD)

خطوات عملية اكتشاف المعرفة (KDD)[١]	
اختيار البيانات Data Selection	يتم اختيار وتحديد البيانات التي يتم عليها تحليل من مجموعة البيانات.
تنظيف وتنقية البيانات Data cleaning	يتم إزالة البيانات الغير مهمة والبيانات المكررة والتي ليست لها صلة بالبيانات التي تم جمعها.
تحويل البيانات Data transformation	يتم دمج البيانات في أشكال محددة ومناسبة عن طريق تلخيص العمليات وتجميعها حتى تكون مهيئة لعملية التنقيب.
التنقيب في البيانات Data Mining	في هذه الخطوة يتم تطبيق تقنيات وأدوات ذكية متنوعة لاستخراج أنماط البيانات.
تقييم الأنماط Pattern evaluation	تعني تحديد بدقة الأنماط التي تمثل المعرفة.
تمثيل المعرفة Knowledge representation	هذه هي المرحلة الأخيرة التي يتم فيها استخدام تقنيات التمثيل المرئي visualizationوالمعرفي لمساعدة المستفيدين على فهم وتفسير معرفة أو النتائج المستخرجة.

جدول رقم (١) يبين خطوات اكتشاف المعرفة (KDD)

[1]) P.Boopathi, L.Ramesh. A Study on Techniques of Data Mining and its Applications. International Journal of Creative Research Thoughts (IJCRT).Volume 6, Issue 1,(March 2018).p 1237.

يوضح الجدول السابق المراحل المختلفة لعملية اكتشاف المعرفة في قواعد البيانات (KDD) والذي يبين أن التنقيب في البيانات جزء من عملية (KDD)، ولأنها جزء أساسي منها يتم استخدام المصطلحين بالتبادل. لذلك قبل أن يتم استخدام أي تقنية في العملية، يجب أن يكون واضحًا ما هو الغرض المقصود من هذه التقنية بعد الاستخدام ، من أجل أداء عملية الاختبار بشكل أفضل. بعد الفهم الكامل لمشكلة النتائج المراد استمرارها، هناك حاجة إلى فهم مفصل لمجموعة البيانات التي سيتم استخدامها في استخراج البيانات. يجب التحقق من تنسيق البيانات وحجم البيانات ونوع البيانات وما إلى ذلك. بالإضافة إلى ذلك، يجب استكمال وتحسين البيانات الغير مهمه والبيانات المفقودة قبل إجراء عمليات التنقيب في البيانات.عادةً ما تأتي مصادر البيانات لعمليات التنقيب في البيانات قواعد بيانات الوسائط المتعددة وقواعد البيانات غير المتجانسة وقواعد البيانات العلائقية. في عملية (DM)، يتم الحصول على النتائج التجريبية النهائية بشكل أساسي من خلال سلسلة من العمليات. لذلك، يصبح الاختيار المعقول والتطبيق الدقيق للنموذج مهمًا للغاية.

٢/٩/١ خطوات عملية التنقيب في البيانات (Data Mining)

	خطوات عملية التنقيب في البيانات (DM)[١]	
تنقيب البيانات	**فهم العمل Understand the business**	يعني فهم طبيعة العمل بالمؤسسة، وذلك من خلال تحويل المشكلة الفعلية التي تتطلب التنقيب في البيانات إلى مشكلة يمكن أن تستخدم نموذجًا

[1]) Wa, Wei. Research and Application of Data Mining Technology in Library Office Information Construction. Journal of Physics: Conference Series. 2020. P4.

		خوارزميًا. استجابة لهذه المشكلة، يتم الحصول على النتيجة المتوقعة. وفي هذه المرحلة يتم جمع البيانات، والتأكد من جودتها.
	فهم البيانات Understanding the data	هو الفهم الكامل لمجموعة البيانات المطلوبة لهذه المشكلة. هذه الخطوة هي أساس عملية التنقيب في البيانات بأكملها. طالما أن اختصاصي البيانات يفهم تمامًا جودة البيانات ونوعها ، يمكنه تطبيق خوارزميات النموذج بشكل أفضل عليها لحل المشكلات.
	معالجة البيانات Data processing	هناك حاجة إلى مساعدة سلسلة من أدوات معالجة البيانات للمعالجة. حيث يتم استخدم وظائف المعالجة المتنوعة لإكمال أعمال تنظيف البيانات وتحويلها وتكميلها وإصلاحها في قاعدة البيانات بأكملها. حيث يتم التخلص من بعض البيانات التي تحتوي على عيوب أو بيانات

		قديمة لا تفيد في الوقت الحالي.
قراءة البيانات وبناء النموذج	**النمذجة Modeling**	هي أكثر الخطوات أهمية في عملية التنقيب في البيانات بأكملها. وهي كيفية اختيار نموذج الخوارزمية ، ما هو نموذج الخوارزمية الذي يجب اختياره، فهو اختيار النموذج الملائم لتمثيل البيانات المكتشفة. كلها قضايا يجب أخذها في الاعتبار بشكل عاجل عند النمذجة.
	تقييم النموذج Model evaluation	تقييم النموذج هو خطوة مهمة بعد اكتمال النمذجة. في هذه الخطوة، يتم الحكم على كل خطوة في النمذجة والأداء العام للنموذج. في الوقت نفسه، يجب مقارنة المشكلات الفعلية والنتائج المتوقعة في خطوة "فهم العمل" لضمان حل المشكلات الأولية بشكل فعال. حيث يتم تقييم النتئج ومراجعة العملية وتحديد الخطوات التالية.
	نشر النموذج Model deployment	ما يتم الحصول عليه في عملية

		تقييم النموذج ليس النتيجة النهائية للنموذج. يجب أيضًا عرض نتائج النموذج للمستخدمين والعملاء ، لذلك هناك حاجة إلى مزيد من أعمال النشر. فمن خلاله يتم التأكد من أن الأنماط الناتجة قابلت الأنماط المطلوبة للتنبأ واتخاذ القرار. وهذا يهدف إلى تسهيل المزيد من الفهم والاستخدام الأفضل لنتائج خوارزميات التنقيب في البيانات.

جدول رقم (٢) يوضح خطوات التنقيب في البيانات

التنقيب في البيانات ومؤسسات المعلومات

إن فكرة التنقيب في البيانات مهمة جدًا في مجال إدارة المكتبات ومؤسسات المعلومات، فهي تقدم نهجًا لاتخاذ قرارات أفضل لأمناء المكتبات لأنها تتحكم في مقدار تدفق البيانات وتنظيمها وتقديمها في شكل معلومة. يمكن أن تكون هذه المعلومات مصدر المعرفة لأمناء المكتبات والتي يمكن أن تساعدهم لاحقًا في اتخاذ قرارات مستنيرة بناءً على معلومات حقيقية ودقيقة. ساعد هذا في تطوير خدمات المكتبة ويقدم طريقة أفضل لاتخاذ القرار الصحيح في الوقت المناسب. إن التنقيب في البيانات هو أحد أكثر التقنيات ملاءمة التي يمكن أن تسهل عملية معلومات البيانات في مكان به تدفق بيانات ضخم مثل المكتبة. من خلال التنقيب في البيانات وتقنياته المتعددة؛ يمكن معالجة البيانات. أيضًا إن استخراج البيانات هو أحد الأدوات

المناسبة للمكتبات الرقمية، حيث أنه من خلال تقنياته مثل التجميع والتصنيفات والربط زادت من مستوى التنظيم داخل المكتبة وأعطت فرصًا أفضل لأمين المكتبات لتقديم خدمات جيدة للمستفيدين.[١] ولأنه يمكن تطبيق تقنيات التنقيب في البيانات في مختلف المجالات حسب ملاءمتها. من أجل توفير خدمات مكتبية تتسم بالكفاءة والفاعلية، ظهرت الحاجة إلى المكتبات الرقمية. ولكن ليست الحل الوحيد لاستكشاف المعلومات المخفية من كمية كبيرة من قاعدة البيانات. وعلى النقيض يمكن القيام بذلك عن طريق تطبيق التنقيب في البيانات على بيانات المكتبة.

تطبيقات التنقيب في البيانات في المكتبات

يوجد العديد من تطبيقات التنقيب في البيانات في المكتبات كالأتي:

١/١٢/١ إدارة مقتنيات المكتبة

يمكن للمكتبات إدارة مقتنياتها بشكل جديد من خلال استخدام تقنية التنقيب في البيانات يذكر منها التالي:

- **معرفة الثغرات في مقتنيات المكتبة:** حيث يمكن الحصول على أنماط معرفية جديدة ودقيقة لم تكن معروفة من قبل وذلك من خلال استخدام خوارزميات التنقيب في البيانات، وبالتالي يتبين لأمناء المكتبات أوجه القصور في المقتنيات وأوجه الزيادة، مما يترتب عليه تقييم المقتنيات في أكثر من جهة، وهذا يساعد وهذا يساعد في مشاركة المصادر مع المكتبات الاخرى وذلك لسد هذه الفجوة لديها ولمساعدة المكتبات الأخرى في مشاركتها في الجوانب الأخرى التي تم تغطيها بشكل جيد.[٢]

[1] Niqresh, Mohammad, "The Influence of Data Mining in Increasing Benefits of Libraries in Jordanian Governmental Universities" Library Philosophy and Practice (e-journal). (2021). P5.

[٢] سيد، أحمد فايز. أدوات التنقيب عن البيانات مفتوحة المصدر دراسة تحليلية تقييمية. مجلة جامعة طيبة للأداب والعلوم الإنسانية ، السنة الخامسة، ع١٠،٢٠١٦م. ح٨١٩.

- **الاقتناء:** إن القانون الثالث لعلم المكتبات هو "كل كتاب قارئه". سيساعد تطبيق التنقيب في البيانات في بيانات المكتبة في معرفة المحتويات المطلوبة الضرورية. سيؤدي ذلك إلى تقليل عمل موظفي المكتبة المتعلق بالاقتناء وكذلك سيضمن الاستخدام الفعال لميزانية المكتبة.[١]
- **مشاركة المصادر:** أصبحت من أهم الأسس التي تقوم عليها المعرفة، خاصة في المكتبات، حيث أن تبادل المعلومات يعد أمرًا مهمًا في مجال المعرفة وخواصها. ولذلك جاءت تقنية التنقيب في البيانات من أجل تسهيل آلية تنظيم البيانات التي تتراكم بكميات كبيرة وتحويلها إلى معلومات قابلة للاستخدام ومن ثم عرضها على المستفيدين من خلال الأجهزة والبرمجيات في المكتبات بما في ذلك البحث والاختيار وغيرها. والأنشطة التي تتم داخل المكتبات.[٢] حيث يمكن الخروج بمجموعة من قواعد الارتباط التي تحدد أكثر المصادر متكررة الحدوث في استخداماتها من داخل مؤسسات أخرى ومدى ارتباطها وعلاقتها وتأثيرها على المصادر الموجودة بالمكتبة.

[1] Sengupta, S,Ibid..
[2] Nguyen, G,et. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review, 52(1), (2019). 77-124.

٢/١٢/١ خدمات المعلومات المتاحة بالمكتبة

- يلعب التنقيب في البيانات دورا كبيرا في تحسين خدمات المعلومات، حيث أن التنقيب في البيانات "ينظف" البيانات الموجودة والمتاحة ويعالجها من أجل ضمان استخراج معلومات عالية الجودة وبالتالي تحقيق أهداف المكتبة في أن يتم تكوين الارتباط بين المستفيدين والمعرفة، بالإضافة إلى توفير الخدمات الإلكترونية من خلال إتاحة مواقع المكتبات للمستفيدين ، وبالتالي تزداد كمية المعلومات ، يتم تصنيف فكرة التنقيب في البيانات وتوزيعها وفقًا للمعلومات التي تحتوي عليها وتقديمها للمستخدمين بطريقة مفهومة ولا لبس فيها.[١]
- بالإضافة إلى ذلك تغير دور أمناء المكتبات الذين أتقنوا المعرفة الجديدة والتكنولوجيا الجديدة ، من مقدم بسيط للمعلومات إلى منظم المعلومات، ومقدم خدمة المعلومات ، ومنتج المعلومات ، ومدير المعلومات. يستخدم أمناء المكتبات تقنية التنقيب في البيانات لجمع ومعالجة وتحليل كمية كبيرة من المعلومات حول سلوك استعارة القراء من خلال شبكات الكمبيوتر، وتحديد مجموعات القراء المحددة، والاهتمامات الفردية، وعادات القراءة، والميول والاحتياجات الخاصة بالمستفيدين.[٢]

٣/١٢/١ تنمية الموارد البشرية

تمتلك المكتبات العديد من الموارد البشرية سواء متخصصة وغير متخصصة، ويمكن تقسيم العاملين الغير متخصصين لأكثر من تخصص، ويترتب عليه وجود العديد من التخصصات والمؤهلات والدرجات العلمية داخل المكتبة الواحدة، وبالتالي يزداد الأمر صعوبة أمام المسئولين ومتخذي القرار في الإلمام بجميع جوانبهم المختلفة. ولكن إذا

[1]) Niqresh, Mohammad, "The Influence of Data Mining in Increasing Benefits of Libraries in Jordanian Governmental Universities" Library Philosophy and Practice (e-journal). (2021). P5

[2]) Z, Xin. The Construction of Smart Library Based on Data Mining Technology. Journal of Physics: Conference Series. 2021. P5.

توافرت قاعدة بيانات تتضمن كافة بيانات العاملين بالمكتبة من حيث (أنواعهم، حالاتهم الإجتماعية، مسكنهم، تخصصاتهم، مؤهلاتهم، هواياتهم، ظروفهم الإقتصادية، خبراتهم... إلى غير ذلك)، مستخدما في ذلك تقنيات التنقيب في البيانات، وبالتالي يمكننا استخراج علاقات وروابط قوية ذات كفاءة عالية بين هذه البيانات واستنباط أنماط معرفية ومعلوماتية تربط كل ذلك ببعضه البعض، فالمعلومات التي تم الحصول عليها لم يكن من الواضح اكتشافها أو الحصول عليها إلا من خلال تقنية التنقيب في البيانات، مما تتيح الفرصة أمام المسئولين ومتخذذي القرار من اكتشاف الموارد البشرية المتاحة لديهم بصورة جديدة ومبتكرة تساعدهم في إعادة توزيعهم داخل الأقسام والأنشطة المختلفة في المكتبة ليس فقط وفق تخصصاتهم ولكن وفق الأنماط المعرفة الذكية الكتشفة، مما يساعد كلا من المكتبة وعامليها على حد سواء.[١]

٤/١٢/١ قواعد بيانات المستفيدين

يتوافر لدى المكتبة العديد من المعلومات التي تتعلق بالمستفيدين ، فيمكن من خلال تقنية التنقيب في البيانات بناء ملفات شخصية لكل مستفيد استنادا إلى تاريخ الاستعارة الخاصة به، وأيضا ربط هذه البيانات مع المقتنيات والأنشطة والخدمات التي تتيحها المكتبة، مما ييسر إحاطته بكل ما هو جديد فيما يخص اهتماماته وتقديم مقترحات بالمواد والموضوعات ذات الصلة بنتيجة الاستعارات مع المستفيدين الآخرين المشتركين معه في نفس مجال الاهتمام مما يساعد على تطوير ملفاتهم الشخصية دون تدخل من المستفيد نفسه. يصبح أمام المسئولين مادة خصبة يمكن من خلالها استخراج معلومات تفيد في كافة الأوجه بالمكتبة.[٢]

---

[١] سيد، أحمد فايز. نفس المرجع . ص ٨٢١.

[٢] قرزيز، أسماء. تنقيب البيانات الضخمة في المكتبات الرقمية: دراسة في المفاهيم وآليات التطبيق. مجلة ببليوفليا لدراسات المكتبات والمعلومات . مج٠٣،ع ١٠. ٢٠٢١. ص ٢٠٥.

النتائج والتوصيات

- أوضحت الدراسة بأهمية التنقيب في البيانات في المكتبات وكيفية تطبيقها في المكتبات ومراكز المعلومات حيث تبين في الآونة الأخيرة في مجال المكتبات والمعلومات، وجود زيادة في عدد الوظائف والتي تهتم بالبيانات بشكل عام، حيث تنوعت المسميات الوظيفية لتشمل:(مختص البيانات، ومكتبي البيانات، ومنسق البيانات، وأرشيفي البيانات، وعالم البيانات)، كما لوحظ زيادة في أعداد التسميات الوظيفية ذات الصلة بالبيانات على سبيل المثال:(مكتبي خدمات البيانات البحثية، ومكتبي إدارة البيانات البحثية، ومكتبي الوعي البياناتي، ومكتبي البيانات البحثية، ومكتبي معالجة البيانات) وغيرها من تسميات ذات صلة.
- يحتاج اختصاصيو البيانات في الوقت الحالي إلى مهارات تقنية لاستخراج البيانات وتحليلها، يحتاج أيضا إلى الجمع بين المهارات والمعرفة المطلوبة لإدارة البيانات، ومعالجة البيانات، والكفاءات لدعم إدارة البيانات.
- لايحتاج اختصاصي البيانات أن يصبح مبرمجاً ، ولكن يجب أن يهتم بمعرفة اللغات ومنطق البرمجة لأجهزة الكمبيوتر. وبالتالي ، سوف يكون على دراية بتشغيل مجموعة من البرامج العلمية المستخدمة لتحويل البيانات.

التوصيات

- زيادة الاهتمام بالدراسات حول تقنيات التنقيب في البيانات Data Mining في المكتبات، والدراسات التي تتعلق بتطبيق آليات التنقيب في البيانات في المكتبات.
- القيام بورش عمل ودورات لأخصائي المكتبات لمعرفة أهمية التنقيب في البيانات وكيفية استخدامها في المكتبات ومراكز المعلومات.

- عمل مقررات لعلم البيانات والتنقيب في البيانات في أقسام المكتبات والمعلومات بصفة عامة وتدريس كيفية استغلال هذه التقنية في التخصص ، وتوافر فرص العمل فيها.
- القيام بدورات وورش عمل لطلاب قسم المكتبات والمعلومات في تقنيات التنقيب في البيانات وذلك بالتعاون مع المتخصصين في المجال لزيادة الوعي بأهمية التنقيب في البيانات في تخصص المكتبات والمعلومات.

قائمة المراجع

المراجع العربية

١. سيد، أحمد فايز. أدوات التنقيب عن البيانات مفتوحة المصدر دراسة تحليلية تقييمية. مجلة جامعة طيبة للأداب والعلوم الإنسانية ، السنة الخامسة، ع٦، ٢٠١٦م.

٢. قرزيز، أسماء. تنقيب البيانات الضخمة في المكتبات الرقمية: دراسة في المفاهيم وآليات التطبيق. مجلة ببليوفليا لدراسات المكتبات والمعلومات . مج٠٣،ع ١٠. ٢٠٢١.

المراجع الأجنبية

1. Amin, M. S., Chiam, Y. K., & Varathan, K. D.. Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, (2019).
2. Bano, Shohar. And Gul, Sumeer, Exploring data mining: facets and emerging trends. Digital Library Perspectives Vol. 37 No. 4, (2021).
3. C ,Chen, Xing, Z., & Liu, Y.. What's spain's paris? mining analogical libraries from q&a discussions. Empirical Software Engineering, 24(3), (2019.
4. Ch, Tiju.. Astudy and survey of big data using data mining techniques.international journal of engineering sciences & research technology. 6(10): (October, 2017).
5. Deepashri, K.S. and Kamath, A. ,"Survey on techniques of data mining and its applications", International Journal of Emerging Research in Management and Technology, Vol. 6 No. 2, (2017).
6. G. Sangeetha Lakshmi, M. Jayashree. Comparative Analysis of Various Tools for Data Mining and Big Data Mining. International Research Journal of Engineering and Technology (IRJET). Volume 06 Issue 04 . (Apr 2019).
7. Ka, Ba. Uma, Mu. An Overview of General Data Mining Tools. International Research Journal of Engineering and Technology (IRJET). 04 ,09 . (Sep -2017).
8. Nguyen, G,et. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artificial Intelligence Review, 52(1), (2019).

9. Niqresh, Mohammad, "The Influence of Data Mining in Increasing Benefits of Libraries in Jordanian Governmental Universities" Library Philosophy and Practice (e-journal). (2021).

10. P.Boopathi, L.Ramesh. A Study on Techniques of Data Mining and its Applications. International Journal of Creative Research Thoughts (IJCRT. Volume 6, Issue 1.(March 2018).

11. R. Shankar, S. Duraisamy. Analysis of Data Mining Tasks, Techniques, Tools, Applications And Trends. IOSR Journal of Computer Engineering (IOSR-JCE). Volume 20, Issue 5, Ver. II (Sep - Oct 2018).

12. S ,Li, et. Problems and changes in digital libraries in the age of big data from the perspective of user services . The Journal of Academic Librarianship, 45(1), (2019).

13. Sen, Shantashree. Applications of data mining in library & information centres: an overview. International Journal of Current Research Vol. 9, Issue, 01, (January2017) .

14. Stancin, Igor and Alan Jović. "An overview and comparison of free Python libraries for data mining and big data analysis." *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (2019).

15. Wa, Wei. Research and Application of Data Mining Technology in Library Office Information Construction. Journal of Physics: Conference Series. 2020.

16. Wang, S., Cao, J., & Yu, P. Deep learning for spatio-temporal data mining: A survey. IEEE Transactions on Knowledge and Data Engineering. (2020).

17. Zh, Xin. The Construction of Smart Library Based on Data Mining Technology. Journal of Physics: Conference Series. (2021).

Abstract:

The study sheds light on the importance of data mining, which arose as a result of the need for a means through which the huge amount of existing data could be preserved, organized, and processed. Therefore, the study addressed the tasks and functions of data mining, and the innovative techniques that have emerged to work on controlling and improving data mining concepts in An attempt to give institutions a more comprehensive view of their data, in addition to their own tools, whether open source or commercial. The researcher also addressed the importance of data mining in institutions and information centers, while clarifying the applications of data mining in libraries and information centers.

The study reached a set of results, the most important of which are the following:

- Recently, in the field of libraries and information, it has been shown that there is an increase in the number of jobs that are concerned with data in general, as job titles have diversified to include: (data specialist, data librarian, data coordinator, data archivist, and data scientist), and an increase

in the number of designations has also been observed. Data-related jobs, for example: (Research Data Services Offices, Research Data Management Offices, Data Awareness Offices, Research Data Offices, Data Processing Offices) and other related designations.

- Data specialists currently need technical skills to extract and analyze data, and also need to combine the skills and knowledge required for data management, data processing, and competencies to support data management.
- A data specialist does not need to become a programmer, but must be interested in knowing the languages and programming logic of computers. Thus, you will be familiar with operating a range of scientific programs used to transform data.

key words:

Data Mining, Machine Learning, Knowledge Discovery in databases, Data Mining applications in libraries, Data Mining tools, Data Science