مجلة خليج العرب
للدراسات الإنسانية والاجتماعية

التنقيب عن البيانات Data Mining والتنقيب في الويب Web Mining ودوره في تعزيز إدارة المعرفة ودعم اتخاذ القرار

Data Mining And Web Mining And Their Role In Enhancing Knowledge Management And Supporting Decision-Making

فتحيه محمد عيسى

Fathia Mohamed Issa

دكتوراه ادارة المعرفة -جامعة الملك عبدالعزيز بجدة - تخصص علم المعلومات - ادارة المعرفة

DOI: https://doi.org/10.64355/agjhss252

مجلة خليج العرب للدراسات الإنسانية والاجتماعية © 2025 / تصدر من مركز السنابل للدراسات والتراث الشعبي
هذه المقالة مفتوحة المصدر موزعة بموجب شروط وأحكام ترخيص مؤسسة المشاع الإبداعي(CC BY-NC-SA)
https://creativecommons.org/licenses/by-nc-sa/4.0/

الملخص:
تهدف هذه الدراسة إلى تسليط الضوء على الدور المحوري لتقنيات **تنقيب البيانات** و**تنقيب الويب** في تعزيز إدارة المعرفة ودعم اتخاذ القرار، خاصة في ظل التزايد الهائل في حجم وتنوع البيانات داخل بيئات الأعمال الحديثة. تستعرض الدراسة المفاهيم الأساسية للتنقيب عن البيانات، ومراحله، ومكوناته الفنية، كما تناقش آليات دمج تقنيات التنقيب مع أنظمة إدارة المعرفة لتعزيز استخراج المعلومات الدقيقة وتحسين جودة القرارات. وركزت الدراسة بشكل خاص على تنقيب الويب بفروعه الثلاثة: تنقيب المحتوى، تنقيب الاستخدام، وتنقيب الهيكل، مع بيان كيف تسهم هذه الفروع في تحليل سلوك المستخدمين، وتصنيف البيانات، وتحسين تجربة التفاعل مع الأنظمة الإلكترونية.

أشارت النتائج إلى أن استخدام تقنيات التنقيب، وخاصة تنقيب الويب يساهم في تحسين الأداء المؤسسي وتطوير الأنظمة التعليمية وفي التسويق الرقمي. كما أوضحت الدراسة أن دمج خوارزميات مثل شجرة القرار والشبكات العصبية وقواعد الارتباط، ضمن منصات مثل WEKA و KNIMEيرفع من كفاءة التحليل ويوسع فرص اكتشاف المعرفة ومن أبرز التحديات التي تناولتها الدراسة التعامل مع البيانات غير المهيكلة وضعف التكامل بين مصادر البيانات، والحاجة إلى تقليل الاعتماد على الخبرة البشرية المكلفة. أوصت الدراسة بضرورة الاستثمار في بناء فرق متعددة التخصصات تجمع بين خبراء المجال ومحللي البيانات بالإضافة تأكيد أهمية اعتماد نماذج مثل KDD و CRISP-DMلتوحيد خطوات التنقيب وتحقيق نتائج قابلة للتكرار.

الكلمات المفتاحية: تنقيب البيانات، تنقيب الويب، إدارة المعرفة.

Abstract:

This study investigates the pivotal role of both data mining and web mining in enhancing knowledge management processes, especially in the context of rapidly expanding and diversifying data environments. It provides a comprehensive overview of data mining concepts, phases, and technical components, while also emphasizing the integration of mining techniques into Knowledge Management Systems (KMS) to support accurate information extraction and informed decision-making. The paper gives particular attention to web mining and its three branches—content, usage, and structure mining—highlighting their utility in analyzing user behavior, organizing data, and optimizing digital experiences.

The findings demonstrate that applying data and web mining significantly enhances organizational performance, educational systems, and marketing strategies. The integration of algorithms such as decision trees, neural networks, and association rules within tools like WEKA and KNIME was shown to improve analytical efficiency and expand knowledge discovery. The study also identifies several key challenges, including handling unstructured data, integrating heterogeneous data sources, and reducing the dependency on costly human expertise. It recommends investing in multidisciplinary teams combining domain experts and data analysts, and adopting standardized frameworks such as KDD and CRISP-DM to ensure methodological consistency and replicable outcomes.

Keywords: Data Mining, Web Mining, Knowledge Management.

المقدمة:

في ظل التطورات المتسارعة في تكنولوجيا المعلومات، أصبحت المعرفة المورد الاستراتيجي الذي لا غنى عنه لتحقيق التميز التنظيمي وتعزيز القدرة التنافسية للمؤسسات. إذ تُمكّن إدارة المعرفة المنظمات من جمع وتخزين وتبادل واستخدام المعرفة بشكل فعّال، مما يساهم في تحسين عمليات اتخاذ القرار ودعم الابتكار المستدام. وتبرز أهمية أدوات تحليل البيانات وتقنيات التنقيب عن البيانات في استخراج المعرفة الكامنة من الكم الهائل من المعلومات المتاحة، حيث تساعد هذه الأدوات في تحويل البيانات إلى طرق قابلة للتنفيذ تدعم تحقيق أهداف المؤسسة وتمنحها ميزة تنافسية في بيئة الأعمال الديناميكية والمتغيرة باستمرار.

تزايد الاهتمام باستخراج البيانات من الويب مثل قواعد البيانات التعليمية والإعلانات وأنظمة المعلومات حيث يشير تعدين الويب إلى تطبيق تقنيات التنقيب عن البيانات لاكتشاف واستخلاص الأنماط والمعرفة من بيانات الويب حيث يدعم عملية صنع القرار وإدارة العمليات في التعليم وتتمثل التحديات العامة في الاستخراج الآلي للمعرفة من البيانات (المقاطع المفسرة من النصوص) ورسم الخرائط لمجموعات البيانات الموجودة، كما توجد العديد من التحديات المتعلقة بتكامل البيانات (Fischer, & Dörpinghaus ,2024, P.53).

كما تهدف تقنية التنقيب عن البيانات إلى استخدام تقنيات معالجة البيانات مثل استخلاصها وتحويلها وتحميلها، للحصول على كميات هائلة من البيانات بأنواع مختلفة بشكل تلقائي وأن انعدام تحليل البيانات يؤدي إلى إهمال المعلومات وعدم معرفة القيّمة الموجودة في مجموعة البيانات الأصلية، لذا أهتم الباحثون والمجتمعات الأكاديمية منذ عقود بجانب تطوير أساليب تحليل البيانات بما في ذلك تحديد مهام التنقيب عن البيانات وتصميم الخوارزميات وتقييم النتائج وغيرها وبذلك يمثل استخراج البيانات مرحلة مهمة في مشروع التنقيب عن البيانات حيث يجب تحديد المهام التي سيتم تنفيذها على البيانات المتاحة، وقد تطوّرت معظم الإنجازات في مجال تحليل البيانات مثل طرق إعداد البيانات وتصميم الخوارزميات ومعايير التقييم إلى نوع من التقنيات الحوسبيّة حيث يمكن استخدام مختلف الأدوات البرمجية مباشرة بالرغم من ذلك لايزال تحديد المهام يعتمد بشكل كامل على الأشخاص بسبب المعرفة والخبرة الغامضة التي يمتلكونها، كما يؤدي ذلك إلى كفاءة منخفضة وتكلفة اتصال عالية.(Wang. et.al. ,2019, P:139537)

ومن خلال استعراضنا لتقنية التنقيب عن البيانات وأهميتها في دعمها لإدارة المعرفة وعملياتها من خلال استخدام وظيفة التنقيب عن البيانات المكونة من (التصنيف والارتباط والتجميع) كدعم لمعالجة البيانات في إدارة المعرفة حيث يمكن أن تختلف الخوارزمية المستخدمة حسب بيانات الشركة التي يتم تحليلها، وتستثمر الكثير من المنظمات من خلال دمج نظم إدارة المعرفة (KMS) مع تقنية التنقيب عن البيانات (DM) لتحسين استراتيجياتها منها التسويق، واستخدام نهج التجميع لاستخراج المعرفة في تطوير(KMS) ودعم إحدى الميزات الرئيسية لـ (KMS) وهي المجموعة المعرفية التي بدورها تساهم في تحسين مخرجات نظم إدارة المعرفة، وكذلك الحصول على المعرفة من قاعدة بيانات المعاملات التاريخية، وتخزين وتنظيم ومشاركة المستندات والمعلومات والمعرفة بشكل منهجي لجميع موظفي المنظمة، وزيادة الميزة التنافسية من خلال تمييزها بين كيفية تسويق منتجاتها المعروضة. (Mardiani et al. ,2019, P :189) .

المنهجية:

اعتمدت هذه الدراسة على المنهج **الوصفي التحليلي**، بهدف تحليل المفاهيم النظرية المرتبطة بتقنيات التنقيب عن البيانات وتنقيب الويب، واستكشاف دورها في دعم إدارة المعرفة وقد تم توظيف هذا المنهج لاستعراض الأدبيات العلمية والدراسات السابقة ذات الصلة، وتحليل الأطر النظرية والنماذج المعتمدة مثل KDD وCRISP-DM، بالإضافة إلى استعراض عدد من الأدوات والتطبيقات البرمجية الشائعة في هذا المجال مثلWEKA ، KNIME، RapidMiner .

كما سعت الدراسة إلى الربط بين الجوانب التقنية والتنظيمية في موضوع إدارة المعرفة، من خلال تحليل العلاقة بين آليات استخراج المعرفة من البيانات، ومدى إمكانية توظيفها في تحسين الأداء المؤسسي واتخاذ القرار كما ركزت على تحليل المحتوى المعرفي المتاح في المصادر العلمية المحكمة، والوثائق التقنية ذات الصلة، بهدف الوصول إلى تصور تكاملي نظري وتطبيقي حول الموضوع.

مشكلة الدراسة:

في ظل النمو المتسارع في حجم البيانات وتنوع مصادرها، تواجه المؤسسات تحديات كبيرة في كيفية الاستفادة من هذه البيانات وتحويلها إلى معرفة مفيدة تدعم اتخاذ القرار وتحسّن الأداء التنظيمي وبالرغم من تطور تقنيات تنقيب البيانات وتنقيب الويب، لكن مازال الاستفادة منها في إدارة المعرفة محدودة أو غير موجهة بشكل جيد. وتكمن المشكلة الرئيسة في غياب إطار تكاملي يربط بين تقنيات التنقيب واستخلاص المعرفة المؤسسية، مما يعيق تحقيق الاستفادة القصوى من الإمكانات التحليلية المتاحة. من هنا، تنبع الحاجة إلى فهم العلاقة بين هذه الأدوات الحديثة للتنقيب البيانات وإدارة المعرفة، واستكشاف مدى قدرتها على توليد معرفة تدعم قرارات فعّالة ومستندة إلى البيانات.

إلى أي مدى تساهم تقنيات تنقيب البيانات وتنقيب الويب في دعم إدارة المعرفة داخل المؤسسات، وما التحديات التي تواجه تكاملها؟

أهداف الدراسة

تهدف هذه الدراسة إلى:

1. توضيح المفاهيم الأساسية لتنقيب البيانات وتنقيب الويب وأدوارهما في المؤسسات.
2. تحليل الأدوات والتقنيات الشائعة المستخدمة في تنقيب البيانات والويب، وتقييم مدى قابليتها للتكامل مع إدارة المعرفة.
3. إبراز المجالات التطبيقية التي تحقق فيها تقنيات التنقيب على تحسين إدارة المعرفة (مثل التعليم، الصحة، التسويق).
4. تقديم توصيات عملية لتعزيز استخدام التنقيب في دعم استراتيجيات إدارة المعرفة داخل المنظمات.

التساؤلات البحثية:

انطلاقًا من مشكلة الدراسة وأهدافها، تتمحور التساؤلات البحثية حول ما يلي:

1. ما مدى إمكانية الاستفادة من تقنيات تنقيب البيانات وتنقيب الويب في دعم إدارة المعرفة؟
2. ما أبرز الأدوات والأساليب المستخدمة في التنقيب عن البيانات والويب التي يمكن دمجها مع إدارة المعرفة؟
3. كيف تسهم عمليات التنقيب في تحسين اتخاذ القرار وتوليد المعرفة التنظيمية؟
4. ما التحديات التي تواجه المؤسسات عند دمج تقنيات التنقيب مع إدارة المعرفة؟
5. ما المجالات التطبيقية التي أظهرت نماذج ناجحة لهذا التكامل؟

أهمية الدراسة:

في ظل التحول الرقمي المتسارع، أصبحت البيانات تمثل المورد الأكثر قيمة في عصرنا الحالي، وأصبح استثمارها أحد أهم مداخل التطوير والتميز المؤسسي ومع هذا التدفق الهائل للمعلومات من مصادر متنوعة، برزت الحاجة إلى أدوات وتقنيات قادرة على تحويل هذه البيانات الخام إلى معرفة قابلة للاستخدام الفعلي في دعم القرار وصنع السياسات. ومن هنا، ظهرت أهمية تنقيب البيانات (Data Mining) وتنقيب الويب (Web Mining) كأدوات تحليلية متقدمة تتيح للمؤسسات الكشف عن الأنماط الخفية، والتنبؤ بالسلوكيات، واستخلاص المعلومات ذات القيمة من كمّ هائل من البيانات، سواء كانت منظمة أو غير مهيكلة.

وفي الوقت ذاته، تمثل إدارة المعرفة (Knowledge Management) أحد التوجهات الإستراتيجية التي تعتمد عليها المنظمات للحفاظ على خبراتها وتطوير قدراتها التنافسية من خلال جمع المعرفة وتوثيقها وتوظيفها. ومن هذا المنطلق، تأتي أهمية دراسة العلاقة بين تقنيات التنقيب ونظم إدارة المعرفة، للوقوف على مدى التكامل الممكن بين الجانبين، واستكشاف أوجه الاستفادة من التحليل الذكي للبيانات في تعزيز رأس المال المعرفي داخل المؤسسات.

الإطار النظري:

1-مفهوم التنقيب عن البيانات وأهميته :

التنقيب عن البيانات هو عملية تحليلية متقدمة تُستخدم لاستخراج أنماط ومعلومات قيمة من مجموعات ضخمة من البيانات باستخدام تقنيات حسابية وإحصائية متطورة. يهدف التنقيب إلى الكشف عن العلاقات والاتجاهات المخفية التي قد لا تكون واضحة بالطرق التقليدية، مما يساعد المؤسسات على اتخاذ قرارات مستنيرة وتحقيق ميزة تنافسية. ويُعتبر التنقيب عن البيانات جزءاً أساسياً من عملية اكتشاف المعرفة في قواعد البيانات، حيث يشمل مراحل متتابعة من معالجة البيانات، اختيار الخوارزميات المناسبة، واستخراج الأنماط ذات الأهمية العملية. (ويكيبيديا العربية، 2024)

التنقيب عن البيانات هو عملية استخراج المعرفة من كميات هائلة من البيانات، باستخدام الخوارزميات الرياضية التي تعتمد على علوم متعددة مثل الإحصاء، علم الآلة، والذكاء الاصطناعي (زعرود، 2015م) **وكما تم تعريفه في دراسة** (Manamperi, 2023, p: 3) أنه تطبيق تكنولوجيا الحوسبة يستخدم التحليل الإحصائي لمجموعة ضخمة من البيانات بهدف تحديد الاتجاهات المهمة ومعرفة والعثور على الأنماط المختلفة.

وتم تعريفه إجرائيا: أن تنقيب البيانات (Data Mining) هو عملية استخراج بيانات قيمة من قواعد بيانات ضخمة حيث يتم استخدام تقنيات تنقيب البيانات لفهم الأنماط والاتجاهات والعلاقات داخل مجموعات بيانات كبيرة للمساعدة في اتخاذ قرارات، إن الويب، كمصدر غني بالبيانات، يوفر فرصة كبيرة لاستخدام تنقيب البيانات لاستكشاف سلوكيات المستخدمين وتحسين الخدمات والمنتجات ، **تكمن أهمية التنقيب عن البيانات** في استخدام تقنيات معالجة البيانات مثل استخلاص البيانات ، وتحويلها، وتحميلها، حتى نحصل على كميات هائلة من البيانات بأنواع مختلفة بشكل تلقائي، وأن انعدام تحليل البيانات يؤدي إلى إهمال المعلومات وعدم معرفة القيّمة الموجودة في مجموعة البيانات الأصلية، لذلك أهتم الباحثون والأكاديميون منذ عقود بجانب تطوير أساليب تحليل البيانات و تحديد مهام التنقيب عن البيانات، وتصميم الخوارزميات وتقييم النتائج وبذلك يمثل استخراج البيانات مرحلة مهمة في مشروع التنقيب عن البيانات حيث يجب تحديد المهام التي سيتم تنفيذها على البيانات المتاحة، ولقد تطوّرت معظم الإنجازات في مجال تحليل البيانات مثل طرق إعداد البيانات وتصميم الخوارزميات ومعايير التقييم إلى نوع من التقنيات الحوسبة حيث يتم استخدام مختلف الأدوات البرمجية مباشرة على الرغم من أن تحديد المهام مازال يعتمد بشكل كامل على

الأشخاص بسبب المعرفة والخبرة البشرية التي اكتسبوها ويؤدي ذلك الى تكلفة عالية في التواصل مع انخفاض الكفاءة فتهدف تقنيات التنقيب عن البيانات الى نمذجة كيفية تحديد الأشخاص لمهام التنقيب منه خلال فهم الأعمال والاستفادة من أساليب اكتشاف المهام وذلك يرفع من الفعالية والجودة ويقلّل من العامل البشري المكثف (Wang, et al. ,2019, P.139537).

2-المكونات التنقيب فى البيانات:

مصادر البيانات: المكان الذي يتم فيه الحصول على البيانات، والتي يمكن أن تكون متنوعة، بما في ذلك الويب كمستودع ضخم للبيانات، أو الملفات النصية، أو جداول البيانات، أو أي مصدر آخر قابل للتطبيق.

قاعدة البيانات أو خادم مستودع البيانات: الخادم الذي يضم جميع البيانات الجاهزة للمعالجة، ويجلب البيانات بناءً على طلبات المستخدم، مما يجعل مجموعات البيانات مخصصة للغاية.

محرك استخراج البيانات: مكون حاسم يحتوي على وحدات مختلفة لمهام مثل الارتباط والتوصيف والتنبؤ والتجميع والتصنيف وما إلى ذلك.

وحدات تقييم الأنماط: تستخدم لقياس أهمية النموذج المبتكر، وذلك باستخدام قيمة عتبة للتقييم.

واجهة المستخدم الرسومية: الواجهة التي تتفاعل مع المستخدم، وتعمل كحلقة وصل حيوية بين المستخدم ونظام استخراج البيانات، مما يبسط تعقيدات استخراج البيانات من أجل تفاعل وفهم سهل الاستخدام.

قاعدة المعرفة: لاكتشاف أنماط مثيرة للاهتمام في البيانات، وتعمل كمرجع لتحديد الأفكار القيمة وتعزيز موثوقية ودقة النتائج النهائية (Sharma,2024,P.1).

3-مكونات هندسة تنقيب البيانات:

التنقيب عن البيانات هو العملية التي يتم فيها استخراج المعلومات التي لم تكن معروفة، والتي تكون مفيدة جدًا، من مجموعة بيانات واسعة جدًا، وإن بنية استخراج البيانات أو بنية تقنيات استخراج البيانات ليست سوى المكونات المختلفة التي تشكل العملية الكاملة لاستخراج البيانات.

1/ مصادر البيانات: يعتبر المكان وجود البيانات أنه مصدر البيانات التي نعمل عليها وتعتبر شبكة الويب مستودع بيانات ضخمة نحصل على البيانات منها وقد تكون البيانات موجودة في ملفات نصية او مستندات جدول او أي مصدر أخر.

2/ قاعدة البيانات أو خادم مستودع البيانات: الخادم هو المكان الذي يحتوي على جميع البيانات الجاهزة للمعالجة. يعمل جلب البيانات حسب طلب المستخدم، وبالتالي، يمكن أن تكون مجموعات البيانات الفعلية شخصية للغاية.

3/ محرك استخراج البيانات: إن مجال استخراج البيانات غير مكتمل بدون ذكر العنصر الأكثر أهمية فيه، والمعروف باسم محرك استخراج البيانات. يحتوي على الكثير من الوحدات التي يمكن استخدامها لأداء مجموعة متنوعة من المهام. يمكن أن تكون المهام التي يمكن تنفيذها هي الارتباط، والتوصيف، والتنبؤ، والتجميع، والتصنيف، وما إلى ذلك.

4/ وحدات لتقييم الأنماط: يتم استخدام هذه الوحدة من الهندسة المعمارية بشكل أساسي لقياس مدى اهتمام النموذج الذي تم تصميمه فعليًا. ولأغراض التقييم، يتم استخدام قيمة العتبة. ويجب ملاحظة أن الوحدة لها رابط مباشر للتفاعل مع محرك استخراج البيانات، الذي يتمثل هدفه الرئيسي في العثور على أنماط مثيرة للاهتمام.

5/ واجهة المستخدم الرسومية: هي ما يتفاعل مع المستخدم. تعمل واجهة المستخدم الرسومية كحلقة وصل تشتد الحاجة إليها بين المستخدم ونظام استخراج البيانات. تتمثل المهمة الرئيسية لواجهة المستخدم الرسومية في إخفاء التعقيدات التي تنطوي على عملية التنقيب عن البيانات بأكملها وتزويد المستخدم بوحدة سهلة الاستخدام والفهم والتي تسمح له بالحصول على إجابة لاستفساراته بطريقة سهلة الفهم.

6/ تعتبر قاعدة كل المعرفة ضرورية لأي بنية لاستخراج البيانات. حيث يتم استخدام قاعدة المعرفة كمنارة إرشادية لنمط النتائج. وقد يحتوي على بيانات مما اختبره المستخدمون. يتفاعل محرك استخراج البيانات مع قاعدة المعرفة لزيادة موثوقية ودقة النتيجة النهائية. حتى وحدة تقييم الأنماط لديها رابط إلى قاعدة المعرفة. يتفاعل مع قاعدة المعرفة على فترات منتظمة للحصول على المدخلات والتحديثات المختلفة منها (Sharma, R,2024, P.1).

4- خصائص التنقيب عن البيانات

إن التنقيب عن البيانات عملية تفاعلية وتكرارية للعثور على أنماط أو نماذج جديدة تكون مثالية ومفيدة وقابلة للفهم في قاعدة بيانات ضخمة، ويتضمن التنقيب عن البيانات البحث عن اتجاهات أو أنماط مرغوبة في قاعدة بيانات كبيرة لمساعدة صانعي القرار مستقبلاً حيث يتم التعرف على هذه الأنماط من قبل أجهزة معينة يمكنها توفير تحليل مفيد وعميق للبيانات يمكن دراسته بشكل أكبر، وربما باستخدام أدوات دعم القرار الأخرى.

ومن أهم خصائص وسمات التنقيب عن البيانات ما يلي:

1/التنقيب عن البيانات يساهم في الكشف عن الشيء المخفي الغير ظاهر، ومعرفة أنماط بيانات معينة لم تكن معروفة من قبل.

2/استخدام التنقيب عن البيانات عادةً للبيانات الكبيرة جداً والضخمة؛ وذلك بهدف جعل النتائج أكثر مصداقية.

3/ إن التنقيب عن البيانات مفيد في عمليات اتخاذ القرارات الحاسمة خاصة في مجال الإستراتيجية.

(Siregar et al ,2023, p 264).

الخصائص الأساسية للتنقيب عن البيانات:

- الكشف عن الأنماط والعلاقات الخفية حيث أن التنقيب عن البيانات يهدف إلى استخراج معلومات جديدة غير معروفة مسبقا من قواعد بيانات ضخمة، من خلال البحث عن أنماط وعلاقات هامة بين البيانات
- استخدام تقنيات الذكاء الاصطناعي والإحصاء تعتمد عملية التنقيب على توظيف أساليب متقدمة من الذكاء الاصطناعي والإحصاء لتحليل البيانات واكتشاف المعرفة المفيدة
- دعم اتخاذ القرار من خلال النتائج المستخرجة من التنقيب التي تساعد المؤسسات وصناع القرار في التركيز على المعلومات الأكثر أهمية، مما يساهم في اتخاذ قرارات سليمة وفعالة
- التعامل مع البيانات الضخمة والمعقدة حيث يتم استخدام التنقيب مع مجموعات بيانات ضخمة ومعقدة، بهدف استخراج القيمة الخفية منها وتسهيل تحليلها
- عملية تكرارية وتفاعلية حيث أن التنقيب عن البيانات ليس عملية خطية، بل يتطلب تكرار المراحل وتحليل النتائج بشكل مستمر للوصول إلى أفضل الأنماط الممكنة

"التنقيب في البيانات هو أسلوب متقدم من أساليب تحليل البيانات يقوم بالبحث والاستكشاف من أجل الحصول على الأنماط غير المعروفة والعلاقات الهامة في قواعد البيانات عن طريق توظيف أساليب الذكاء الاصطناعي والإحصاء وتقنيات قواعد البيانات المتقدمة، أي أنه مظلة كبيرة ومفهوم واسع يندرج تحته العديد من المهام والأدوات من أجل الحصول على المعلومات المفيدة والداعمة لاتخاذ القرار السليم (العازمي، 2023، ص. 4) .

5- عمليات التنقيب ما يلي:

1. تنظيف البيانات من خلال إزالة الضوضاء وعدم التناسق البيانات، والجمع بين مصادر بيانات متعددة.

2. تكامل البيانات وذلك بالجمع بين مصادر بيانات متعددة.

3. اختيار البيانات ذات الصلة تم استرجاعها من قاعدة البيانات المطلوبة في التحليل.

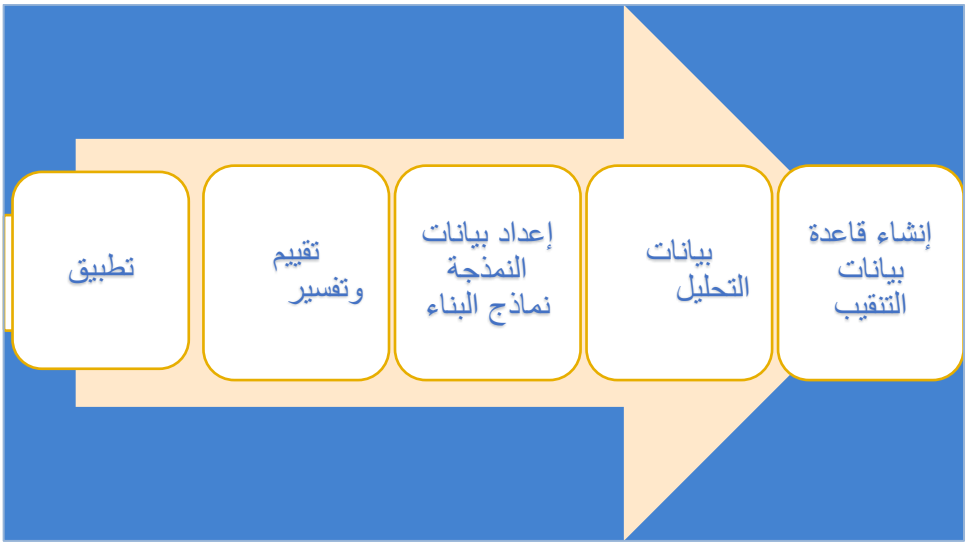
4. تحويل البيانات إلى النماذج المناسبة لأداء استخراج البيانات.

5. استخراج البيانات من خلال اختيار خوارزمية استخراج البيانات ينطبق ذلك على النمط الموجود في البيانات.

6. تقييم النمط بتحديد المثير للاهتمام الأنماط التي تمثل المعرفة.

7. عرض المعرفة: ترجمة الأنماط المفيدة إلى مصطلحات يمكن للبشر أن يفهموها (Alejandrino,2021,P.2691)

عمليات التنقيب في البيانات:



التنقيب في البيانات من المصدر (Yu & Li ,2021,P.3)

6- مهام التنقيب عن البيانات التي تدعم عمليات التنقيب ومنها:

يوضح الجدول أدناه ، جدول (3) مقارنة بين تقنيات استخراج البيانات مزايا وعيوب مهام التنقيب عن البيانات، حيث أن يساعد فريق العمل ب اختيار المهام حسب طبيعة البيانات وأيضا حسب المجال التي يتم تنقيب البيانات منه ، إن تقنية التنقيب عن البيانات تعتمد على مجموعة واسعة من المتطلبات منها التقنيات المستخدمة والمهارات البشرية لتحليل البيانات واستخراج المعلومات المفيدة منها حيث يتم جمع مجموعة كبيرة من البيانات حول الموارد المراد التنقيب عنها، هناك العديد من المتطلبات للتنقيب عن البيانات منها وجود فريق عمل مهمته التنقيب عن البيانات وتخزين البيانات وإدارتها، حيث يتكون فريق العمل من ثلاثة أنواع رئيسية من الخبراء:

خبراء المجال- محلّلو البيانات- متخصصون في تكنولوجيا المعلومات.

.(Kara, M., Oktay Fırat, S. and Ghadge, A., 2019.P12)

7- مراحل تقنية التنقيب عن البيانات

تتضمن مراحل تقنية تنقيب البيانات الخطوات التالية:

- يتم اختيار البيانات من مصادر بيانات مختلفة ومتنوعة مطلوبة لعملية تنقيب البيانات.
- يتم المعالجة المسبقة التي تتضمن البحث عن القيم الغير مألوفة أو القيم المفقودة.
- يتم تحويل البيانات إلى تنسيق مشترك للمعالجة، وتستخدم تقنيات التقليل من حجم البيانات وتقنيات تقليل الأبعاد لتقليل عدد القيم الغير مهمة.
- يتم تنقيب عن البيانات وذلك لتوليد الأنماط من مجموعة البيانات المعطاة.
- يتم التفسير/التقييم حيث يتطلب تقديم نتائج تنقيب البيانات للمستخدم.

.(Dol &Jawandhiya ,2022,p: 381-385)

7-مجالات استخدام التنقيب في البيانات:

1- يتم استخدام تنقيب البيانات في الأعمال التجارية لتحليل البيانات واستخراج المعارف التي تساعد في اتخاذ القرارات الاستراتيجية وتحسين العمليات.

2- يمثل تنقيب البيانات مهم في الصناعة المالية من خلال تحليل البيانات المالية وتوجيه القرارات الاستثمارية وإدارة المخاطر.

3- يستخدم تنقيب البيانات في مجال التسويق لفهم سلوك العملاء وتحليل البيانات الخاصة بالعملاء لتحسين استراتيجيات التسويق وزيادة كفاءة الحملات التسويقية.

4- يساهم تنقيب البيانات في تحسين الرعاية الصحية من خلال تحليل البيانات الطبية وتحديد الاتجاهات والأنماط لتحسين تشخيص الأمراض وتحسين جودة الرعاية الصحية.

5-تنقيب البيانات ضروري في إدارة المعرفة من خلال تحليل البيانات واستخلاص الرؤى القيمة التي تساعد في تبادل المعرفة داخل المؤسسة وتحسين عمليات اتخاذ القرارات(Shettigar, et al. ,2022, p: 2286, p :2290-p 2291).

أيضا من مجالات استخدام التنقيب عن البيانات كما يلي :

- يستخدم التنقيب عن البيانات لتحليل سلوك العملاء، تحسين استراتيجيات التسويق، التنبؤ بالطلب، وتحديد فرص البيع المتقاطع (Cross-selling)وزيادة ولاء العملاء.
- يُستخدم في تقييم المخاطر، كشف الاحتيال المالي، التنبؤ بالأداء المالي، وتحليل القروض، مما يساعد في اتخاذ قرارات ائتمانية سليمة.
- تحليل بيانات المرضى لتحسين التشخيص، التنبؤ بالأمراض، تحسين جودة الرعاية الصحية، وإدارة الموارد الطبية بكفاءة.
- التصنيع وسلسلة التوريد وتحسين عمليات الإنتاج، التنبؤ بصيانة المعدات، وتحليل جودة المنتجات لتقليل التكاليف وزيادة الكفاءة.
- الاتصالات وتكنولوجيا المعلومات وتحسين خدمات العملاء، تقليل التسرب، وتحليل استخدام الشبكات لتقديم عروض مخصصة.
- التعليم وتقييم أداء الطلاب، تخصيص المناهج التعليمية، وتحليل بيانات التعلم لتحسين النتائج التعليمية Han, Kamber, & Pei . (2012)

8- متطلبات التنقيب عن البيانات:

إن تقنية التنقيب عن البيانات تعتمد على مجموعة واسعة من المتطلبات التقنية والمهارات لتحليل البيانات واستخراج المعلومات المفيدة منها حيث تقوم بجمع مجموعة كبيرة من البيانات حول الموارد المراد التنقيب عنها، حيث انه هناك العديد من المتطلبات للتنقيب عن البيانات منها وجود فريق عمل مهمته التنقيب عن البيانات وتخزين البيانات وإدارتها، حيث يتكون فريق العمل من ثلاثة أنواع رئيسية من الخبراء:

- خبراء المجال هم الأشخاص الذين يمتلكون خبرة ومعرفة داخل مجال التطبيق المحدد، ويتم اختيارهم لتحديد عوامل الخطر حيث يقع على عاتقهم مسؤولية تحليل وتقييم الأوضاع المحتملة للمخاطر واتخاذ القرارات المناسبة للتعامل معها.

- خبراء تنقيب البيانات أو محلّلو البيانات هم الذين يستخدمون أدوات وتقنيات التنقيب عن البيانات في تحليل البيانات، ويتمتعون بمعرفة تقنية واسعة في تحليل البيانات وإدارتها، وتطبيق أساليب مختلفة لاستخراج القيمة من البيانات، ولذلك لابد أن يتمتع خبير البيانات بمستوى معين من المعرفة ليتفاعل مع أعضاء الفريق.

- متخصصون في تكنولوجيا المعلومات او خبراء تكنولوجيا المعلومات يعتبروا من فريق محلّلي البيانات يقومون بتكامل البيانات وإدارتها، وتطوير بنية البيانات والبنية التحتية لتكنولوجيا المعلومات، والوصول للبيانات وتطوير مستودع البيانات وتحليل البيانات بشكل يعمل على تحقيق الأهداف المحددة للمشروع(Er Kara, M., Oktay Fırat, S. and Ghadge, A. ,2019: (P 12

وأيضا من متطلبات التنقيب عن البيانات لابد من استخدام تقنيات تدعم عمليات التنقيب ومنها:

التجميع:(Clustering)

التجميع هو تقسيم المعلومات إلى مجموعات من الكائنات المترابطة. حيث يصف البيانات بواسطة عدد قليل من المجموعات مما يؤدي إلى فقدان بعض التفاصيل الدقيقة، يعتبر التجميع من ناحية عملية أساسيًا في تطبيقات التنقيب عن البيانات ومن بعض أمثلتها: استكشاف البيانات العلمية، تعدين النصوص، استرجاع المعلومات، تطبيقات قواعد البيانات المكانية، إدارة علاقات العملاء، التحليلات الويب، البيولوجيا الحاسوبية، التشخيص الطبي.

الانحدار:(Regression)

تحليل الانحدار في عملية التنقيب عن البيانات يستخدم لتحديد وتحليل العلاقة بين المتغيرات بسبب وجود عامل آخر. حيث يتم تحديد احتمالية متغير معين. بشكل أساسي، يعتبر الانحدار شكل من أشكال التخطيط والنمذجة. فقد نستخدمه للتنبؤ ببعض التكاليف، استنادا على عوامل أخرى مثل التوافر، الطلب الاستهلاكي، والمنافسة.

قواعد الارتباط:(Association Rules)

إن هذه التقنية تساعد في اكتشاف رابط بين عنصرين أو أكثر. تجد نمطًا خفيًا في مجموعة البيانات. قواعد الارتباط هي عبارات شرطية تساعد في إظهار احتمالية التفاعلات بين عناصر البيانات داخل مجموعات بيانات كبيرة في أنواع مختلفة من قواعد البيانات. يتم استخدام تعدين قواعد الارتباط على نطاق واسع لدعم الترابطات في البيانات التجارية أو مجموعات البيانات الطبية.

القيم الشاذة:(Outlier Detection)

هذه التقنية في التنقيب عن البيانات تتعلق بمراقبة عناصر البيانات في مجموعة البيانات التي لا تتطابق مع نمط متوقع أو سلوك متوقع. يمكن استخدام هذه التقنية في مجالات متنوعة مثل الكشف عن التسللات، كشف الاحتيال، وغيرها. تعرف أيضًا بتحليل القيم الشاذة أو التعدين للقيم الشاذة. القيمة الشاذة هي نقطة بيانات تختلف كثيرًا عن بقية مجموعة البيانات. معظم مجموعات البيانات في العالم الحقيقي تحتوي على قيم شاذة. حيث أن كشف القيم الشاذة مهمة في مجال التنقيب عن البيانات. كشف القيم الشاذة قيم في العديد من المجالات ومن بعض أمثلتها: تحديد انقطاعات الشبكة، كشف الاحتيال ببطاقات الائتمان، وكشف القيم الشاذة في بيانات شبكة الاستشعار اللاسلكية.

الأنماط المتسلسلة(Sequential Patterns)

النمط المتسلسل هو تقنية التنقيب عن البيانات المتخصصة لتقييم البيانات المتسلسلة لاكتشاف الأنماط المتسلسلة. وتتضمن هذه التقنية العثور على تسلسلات فرعية ضمن مجموعة من التسلسلات، حيث يمكن قياس أهمية التسلسل حسب المعايير مختلفة مثل الطول، تواتر الحدوث، وغيرها. بمعنى آخر، تساعد هذه التقنية في التنقيب عن البيانات على اكتشاف أو التعرف على الأنماط المشابهة في بيانات المعاملات على مر الزمن.

التنبؤ(Prediction) التنبؤ يستخدم مجموعة من تقنيات التنقيب عن البيانات الأخرى مثل الاتجاهات، التجميع، التصنيف، يقوم بتحليل الأحداث أو الحالات السابقة بالتسلسل الصحيح للتنبؤ بأحداث مستقبلية (Jaiswal ,2024, P :1).

وأيضا هناك تقنيات تنقيب البيانات حسب ما تم ذكرها في دراسة (Beulah , 2019, P:516) منها ما يلي: التصنيف (Classification)

يستخدم لتصنيف البيانات إلى فئات محددة حسب سماتها المشتركة.

يساعد في فهم التنبؤ بالتصرفات المستقبلية بناءً على السجلات السابقة.

التجميع (Clustering):

يقوم بتجميع البيانات المماثلة حسب سماتها المشتركة دون الحاجة إلى تصنيف مسبق.

يمكن استخدامه لاكتشاف الأنماط والمجموعات الطبيعية في البيانات.

شجرة القرار(Decision Tree) :

هي نموذج يستخدم للتنبؤ ويمكن تصورها كشجرة كل فرع في الشجرة يدل على شرط معين، والأوراق تدل على النتيجة إذا تم تلبية الشرط. يتم تقديم كل فرع بسؤال أو شرط يحتوي على إجابتين أو أكثر. كل إجابة قد تقود إلى سؤال آخر أو شرط آخر. تقوم شجرة القرار بتصنيف البيانات حسب للشروط دون فقدان أي بيانات وتساعد في اتخاذ القرارات. وتعمل شجرة القرار كأداة دعم للقرارات. تشبه الرسم البياني حيث يمثل كل عقدة داخلية "اختبارًا" على صفة ما. تدل المسارات من الجذر إلى الورقة على قواعد التصنيف.

الشبكات العصبية:(Neural Network)

هي نظم لاكتشاف أنماط مختلفة وإجراء تنبؤات. يتم تقديم مجموعة من المدخلات إلى الشبكة العصبية، والتي تستخدم بعد ذلك للتنبؤ بواحد أو أكثر من المخرجات. تعطي الشبكات العصبية نتائج عددية حيث يتم استخدام الشبكات العصبية على نطاق واسع في العديد من التطبيقات مثل كشف الاحتيال، التنبؤ بردود فعل العملاء، فهم الصور، كما أنها تساعد في تحليل القيم الشاذة أو في التجميع. تعتبر الشبكات العصبية من الأدوات القوية في مجال التنقيب عن البيانات والتعلم الآلي لقدرتها على معالجة وتحليل كميات كبيرة من البيانات بطريقة تحاكي آلية عمل الدماغ البشري.

ألات دعم المتجهات: يتم تعلم التقنية تحت الإشراف، يساعد في تقليل المخاطر وتقليل التصنيف خطأ، جزء من المصنفات الخطية ويمكن أن يكون امتدادا للإدراك تقنية.

تعلم قواعد الارتباط(Association Rule Learning)

هو تقنية تبحث عن أنماط متشابهة ضمن قاعدة بيانات معينة وتجد قواعد قوية ضمن تلك القاعدة (أي التحديد الدقيق للملاحظات التي تحدث بتكرار). تستخدم قواعد الارتباط قيمة دعم دنيا وقيمة ثقة دنيا يحددهما المستخدم. هذه القيم تحدد مدى قوة وأهمية العلاقات، مما يسمح بتحديد الأنماط التي لها تأثير أكثر والاستمرار في حدوثها ضمن البيانات المحللة (Beulah, 2019, P:514).

يوضح الجدول التالي مزايا وعيوب بعض تقنيات التنقيب عن البيانات:

التقنية	المزايا	العيوب
التصنيف (Classification)	يصنف البيانات حسب سماتها المشتركة، فهم التنبؤ بالمستقبل	حساسة للبنية المحلية للبيانات يتطلب ذاكرة كبيرة.
التجميع (Clustering):	يقوم بتجميع البيانات المشتركة تصنيف مسبق. يكتشف المجموعات الطبيعية في البيانات	تم تنفيذ الدمج أو الانقسام ولا يمكن التراجع عنه أو مشتق.
شجرة القرار	سهولة الفهم والتفسير، القدرة على التعامل مع كل من البيانات العددية والفئوية، قدرتها على العمل مع مجموعات البيانات الكبيرة، سهولة الاستخدام	تكون بعض الحسابات معقدة، غير مناسبة للمهام التي تتطلب تنبؤات مستمرة
الشبكات العصبية	قدرة التعلم الجيدة، سرعة جيدة، واسعة النطاق في الاستخدام	تعامل فقط مع البيانات العددية، قد تواجه مشكلة الحلول المحلية المثالية
آلات دعم المتجهات	تنتج تصنيفات دقيقة للغاية، تتعامل مع الضوضاء	سرعة بطيئة لكل من التدريب والاختبار، قد تحتوي على بيانات متقطعة
تعلم قواعد الارتباط	سهولة الفهم والتفسير، القدرة على التعامل مع مختلف أنواع البيانات، القدرة على العمل مع مجموعات البيانات الكبيرة، سهولة الاستخدام يساعد في عملية اتخاذ القرارات فيرفع من قيمة التقنية بشكل كبير تسمح باكتشاف أنماط متسلسلة في مجموعات البيانات، يسهل التنبؤ بالأحداث والسلوكيات المستقبلية حسب الأنماط المكتشفة. تساعد في اكتشاف العلاقات والارتباطات بين مجموعات متنوعة من البيانات. يمكن استخدامها في التسويق لفهم سلوك المستهلكين وتوجيه الحملات الإعلانية بفعالية.	تكون بعض الحسابات معقدة، غير مناسبة للمهام التي تتطلب تنبؤات مستمرة
التعلم الآلي	قدرة التعلم الجيدة، سرعة جيدة، واسعة النطاق في الاستخدام	تعامل فقط مع البيانات العددية، قد تواجه مشكلة الحلول المحلية المثالية
كشف الشذوذ (Anomaly Detection)	يساعد في اكتشاف البيانات غير المتوقعة أو الشاذة في مجموعات البيانات. يمكن استخدامه في الكشف عن الاحتيال أو الأخطاء في البيانات.	

تعدين الأنماط المتسلسلة	**يوجد نوعان من تعدين الأنماط المتسلسلة وهما تعدين السلاسل وتعدين مجموعات العناصر. في تعدين مجموعات العناصر، يتعرف الخوارزم على مجموعات من الملاحظات التي تحدث بشكل متكرر باستخدام قواعد مختلفة.**	**قد يكون من الصعب الحصول على نتائج دقيقة، قد تكون عملية التعدين النصي بطيئة**
تحليل الانحدار (Regression Analysis)	**-إيجاد ملاحظات متشابهة من أنماط مختلفة. -يقوم بتحليل العلاقة بين متغيرين مختلفين ويكتشف كيف تتغير قيمة المتغير التابع مع تغير قيمة المتغير المستقل- يستخدم في المجالات العلمية والتجارية لتقديم توقعات دقيقة ولفهم العلاقات بين المتغيرات. - تطبيقاته تقييم الأثر الاقتصادي، تحديد الاتجاهات السوقية، البحوث الطبية وفي كثير من المجالات الأخرى. حيث يتطلب فهم العلاقات بين الظواهر المختلفة**	**عدم استقر النماذج المستخدمة في التحليل الإحصائي أو التنبؤ غير مستقرة ويمكن أن تتغير بشكل كبير استجابةً لتغيرات بسيطة في البيانات. هذا يمكن أن يؤدي إلى نتائج غير موثوقة ويجعل من الصعب الحصول على تنبؤات دقيقة أو تحليلات ثابتة.**

الشكل (1) مقارنة بين تقنيات استخراج البيانات الجدول من إعداد الطالبة من مصدر

Beulah, 2019, P:516)

9- التحديات في تنقيب البيانات:

- صعوبة التعامل مع البيانات غير المنظمة التي لا تتبع نمطًا ثابتًا، مما يصعّب معالجتها وتحليلها.
- عدم انسجام البيانات المجمعة من مصادر متعددة يؤدي إلى مشاكل في توحيدها ضمن بيئة تحليل واحدة.
- الكفاءة العالية المطلوبة من المحللين ترفع التكلفة وتعتمد على المهارات الفردية.
- التحديات التقنية في تنقيب الويب
- مثل الطبيعة الديناميكية للويب، والغموض، وعدم تجانس مصادر البيانات.
- خاصةً عند جمع وتحليل بيانات حساسة من مصادر مثل السجلات الطبية أو التفاعلات عبر الإنترنت.
- البيانات المفقودة أو المتكررة أو غير الدقيقة قد تؤثر في نتائج التنقيب.

وجدت الباحثة من أبرز التحديات التي تناولتها الدراسة هي التعامل مع البيانات غير المهيكلة، ضعف التكامل بين مصادر البيانات، والحاجة إلى تقليل الاعتماد على الخبرة البشرية المكلفة.

تحديات التكامل بين تقنيات التنقيب وإدارة المعرفة:

على الرغم من الإمكانات الكبيرة التي توفرها تقنيات تنقيب البيانات وتنقيب الويب في دعم نظم إدارة المعرفة، إلا أن هذا التكامل يواجه العديد من التحديات التي قد تعيق تحقيق أقصى استفادة معرفية داخل المؤسسات. كما تتمثل التحديات في **طبيعة البيانات غير المهيكلة**، حيث تمثل نسبة كبيرة من البيانات المتاحة اليوم، مثل النصوص الحرة، رسائل البريد الإلكتروني، محتوى مواقع الويب، مما يجعل معالجتها وتحليلها أكثر تعقيدًا مقارنة بالبيانات المنظمة بالإضافة الى **ضعف التكامل بين مصادر البيانات**، حيث يتم تجميع البيانات من أنظمة متعددة غير مترابطة (مثل قواعد بيانات منفصلة، أو مصادر خارجية)، مما يؤدي إلى تكرار، أو فقدان، أو تضارب في المعلومات، ويُضعف دقة استخراج المعرفة.

وأيضا الاعتماد **المرتفع على الخبرة البشرية** في تصميم خوارزميات التنقيب وتفسير نتائجها تحديًا من حيث التكلفة والوقت، خاصةً في المؤسسات التي تفتقر إلى كفاءات متخصصة في علم البيانات.

كما تواجه عمليات **تنقيب الويب تحديات تقنية فريدة** مثل التغير المستمر في بنية المواقع، والغموض الدلالي للمحتوى، وصعوبة تتبع سلوك المستخدم عبر منصات متعددة، مما يُضعف من فعالية نماذج الاستخدام والاستخلاص ، وبرزت **مخاوف أمنية وأخلاقية** مرتبطة بجمع وتحليل بيانات المستخدمين، خاصة عند التعامل مع بيانات حساسة (مثل التعليم أو الصحة)، ما يتطلب توازنًا دقيقًا بين استخراج المعرفة والحفاظ على الخصوصية، بالإضافة أن **جودة البيانات وسرعة الاستجابة** تؤثر في كفاءة نظم التنقيب، حيث أن وجود بيانات ناقصة أو مكررة أو قديمة قد يؤدي إلى نتائج تحليلية غير موثوقة، بينما يؤدي بطء استخراج النتائج إلى إضعاف قيمتها في دعم اتخاذ القرار بالزمن الفعلي ، كل هذه التحديات دعت الحاجة إلى تطوير حلول ذكية وشاملة تدمج بين الجانب التقني والبشري، وتدعم التكامل السلس بين أدوات التنقيب ومكونات إدارة المعرفة.

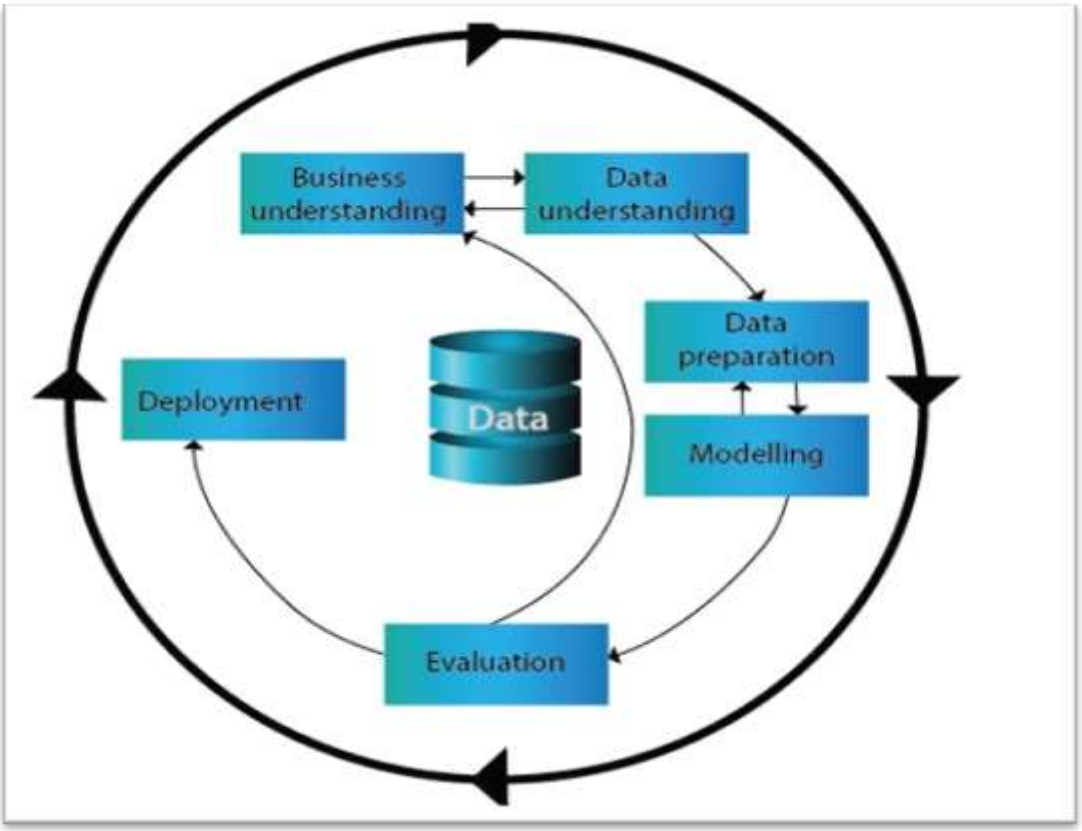
10. آلية عمل تقنية التنقيب عن البيانات

يتم تطبيق نموذج عملية تنقيب البيانات لاستخراج وتنقيب البيانات الكبيرة لإنشاء المعرفة، وإن أكثر ثلاثة نماذج معروفة لعملية تنقيب البيانات هي نموذج العمليات القياسي عبر الصناعات لتنقيب البيانات (CRISP-DM)، ونموذج SEMMA (Sample, Explore, Modify, Model and Assess)، ونموذج اكتشاف المعرفة في قواعد البيانات (KDD)

CRISP-DM-1 نموذج تم تطويره في منتصف التسعينيات من قبل أربع منظمات بارزة ، ومعترف به من قبل الممارسين الصناعيين كنموذج قياسي بسبب قابليته للتطبيق عبر مختلف الصناعات، ويتكون من ست خطوات تكرارية تسمح للمستخدم النهائي بالتحرك إلى الأمام والخلف بمرونة.

SEMMA-2 :أداة منهجية تم تطويرها من قبل معهد SAS للتحليلات الإحصائية والمتحللة حيث يتم أخذ عينات من البيانات، استكشافها، تعديلها، نمذجتها، وتقييمها.

تتكون العملية القياسية للتنقيب عن البيانات من ست مراحل مصممة موضحة في الشكل الآتي:



الشكل (2) يوضح: العملية القياسية عبر الصناعة لاستخراج البيانات (CRISP-DM)، Jaiswal ,2024, P1

KDD-3 هي العملية الشاملة لاكتشاف المعرفة في البيانات حيث تستخدم طرق تنقيب البيانات مثل الإحصاءات وتعلم الآلة والذكاء الاصطناعي وتصور البيانات، وتتكون من تسع خطوات تكرارية وتفاعلية حيث يشير التكرار ، أن المستخدم النهائي يتطلب منه مراجعة الخطوات السابقة والتفاعل يعني أن المستخدم يستطيع المشاركة في كل خطوة.(Md Salamat, et al, 2020, P :4-5) .

11- اكتشاف المعرفة في قواعد البيانات (KDD):

عملية KDD تتمتع بجوانب الذي لا يمكن فيه تقديم صيغة واحدة أو إجراء تصنيف علمي كامل للقرارات الصحيحة لكل خطوة ونوع تطبيق لذلك من المهم فهم العملية والمتطلبات المختلفة والإمكانيات في كل مرحلة تبدأ العملية بتحديد أهداف اكتشاف المعرفة وتنتهي بتنفيذ المعرفة التي اكتشفناها. في هذه النقطة، يتم إغلاق الحلقة، ويبدأ تنقيب البيانات النشط وبالتالي تحتاج إلى إجراء تغييرات في مجال التطبيق مثل تقديم ميزات متنوعة لمستخدمي الهواتف الخلوية من أجل تقليل التحول. يتم بذلك إغلاق الحلقة، ويتم قياس التأثيرات ثم على مستودعات البيانات الجديدة، ويتم عملية اكتشاف المعرفة مرة أخرى.

اكتشاف المعرفة في قواعد البيانات : هو عملية تحليلية تهدف إلى استخراج المعلومات المفيدة والأنماط الغير واضحة من مجموعات البيانات الكبيرة، وتشمل هذه العملية عدة خطوات رئيسية كما وردت في دراسة (زعرود، 2015 م) :

1/اكتشاف البيانات (Data Discovery) وذلك بتحديد واستخراج البيانات ذات الصلة للتحليل.

2/تنظيف البيانات (Data Cleaning) من خلال إزالة الأخطاء والبيانات الضوضائية (Noise) من المجموعة لتحسين جودة البيانات.

3/دمج البيانات (Data Integration) من مصادر متعددة لتوفير منظور شامل.

4/اختيار البيانات (Data Selection) المناسبة للتحليل من المجموعة الكبيرة.

5/تحويل البيانات (Data Transformation) إلى صيغة مناسبة للتحليل.

6 /تنقيب البيانات (Data Mining) بتطبيق الأساليب الإحصائية وخوارزميات التعلم الآلي لاستخراج الأنماط.

7/ تقييم الأنماط (Pattern Evaluation) من خلال تحديد مدى فائدة وموثوقية الأنماط المكتشفة.

8/تقديم المعرفة (Knowledge Presentation) بعرض الأنماط والمعرفة المكتشفة بطريقة يمكن فهمها واستخدامها في صنع القرار (زعرود، 2015).

فيما يلي وصف موجز لعملية اكتشاف المعرفة المكونة من تسع خطوات ابتداءً بخطوة إدارية:

1- بناء فهم لمجال التطبيق حيث يتضمن تطوير فهم لأهداف المشروع والبيئة التي سيتم فيها عملية اكتشاف المعرفة وتحديد أهداف المستخدم النهائي والمعرفة السابقة ذات الصلة.

2-اختيار وإنشاء مجموعة بيانات ويشمل تحديد البيانات المتاحة والمهمة لعملية اكتشاف المعرفة، ودمج جميع البيانات المهمة في مجموعة واحدة.

3-التحضير والتنقية: يتم من خلال تحسين موثوقية البيانات، وتنظيف البيانات من الكميات المفقودة والضوضاء.

4-تحويل البيانات ويتضمن تطوير بيانات مناسبة لعملية تعدين البيانات، وتقليل الأبعاد واستخراج الميزات كذلك تحويل السمات العددية.

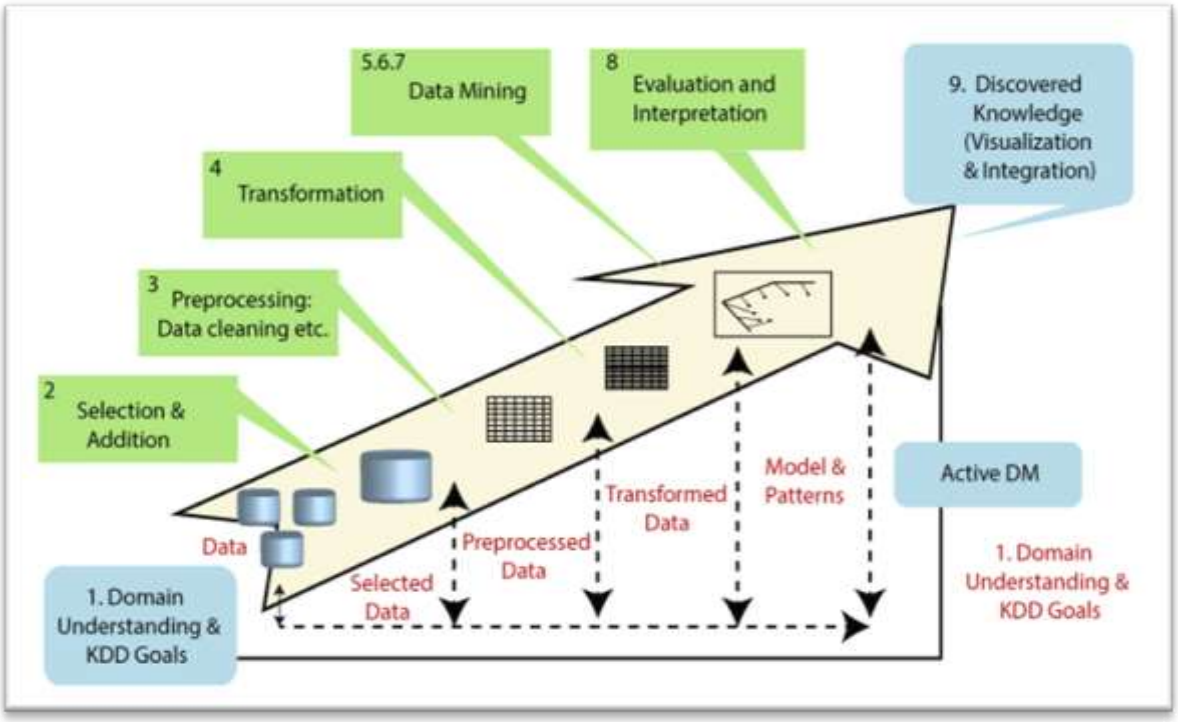
5-التنبؤ والوصف حيث **يتم** من خلال اختيار نوع تعدين البيانات المناسب مثل (تصنيف، تجميع، إلخ) والتركيز على التنبؤ والوصف حسب الأهداف المحددة.

6-اختيار خوارزمية تعدين البيانات ويشمل اختيار الخوارزمية المناسبة للبحث عن الأنماط.

7-استخدام خوارزمية تعدين البيانات ويتم تنفيذ الخوارزمية وضبط معلماتها حتى الحصول على نتائج مرضية.

8-التقييم ويكون ذلك عبر تقييم وتفسير الأنماط المكتشفة وتوثيقها للاستخدام اللاحق.

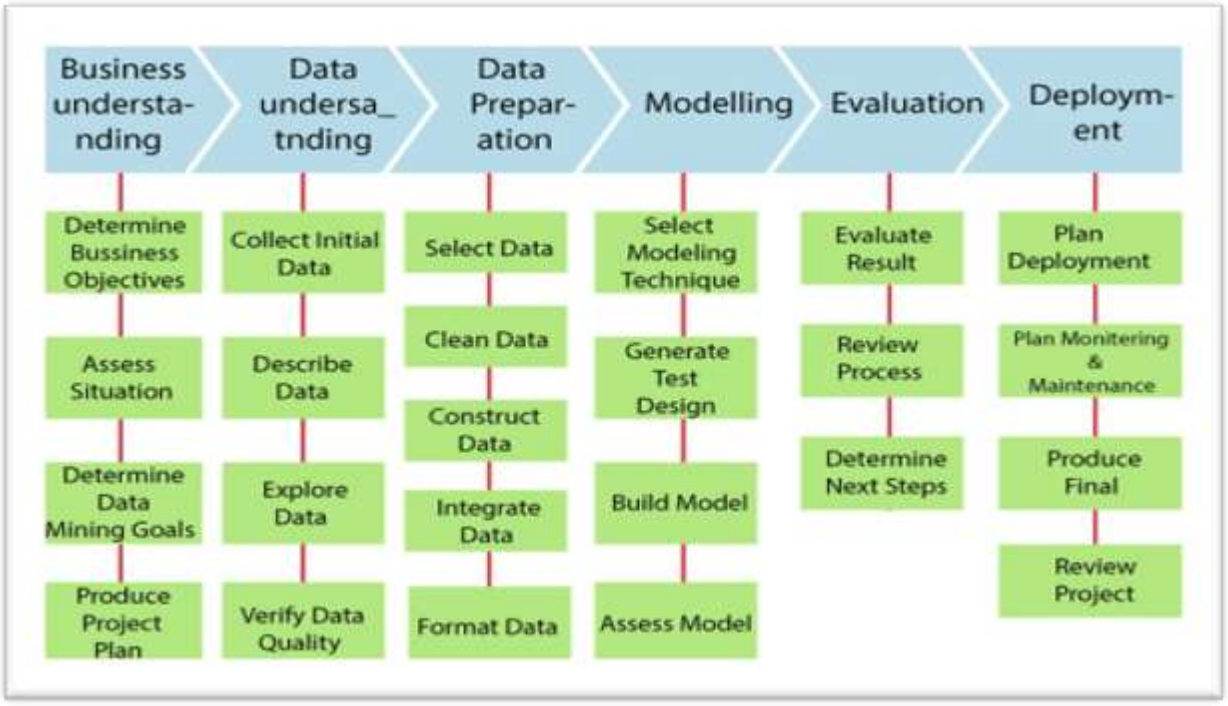
9-استخدام المعرفة المكتشفة: يتم دمج المعرفة في أنظمة أخرى وقياس تأثيرها، والتعامل مع التحديات المتعلقة بتحويل البيانات وديناميكيتها.



الشكل (3) مخطط عملية تنقيب البيانات (KDD Process in Data Mining) (Jaiswal ,2024, P :1)

استفادت الكثير من القطاعات المختلفة من تقنيات تقنيب في البيانات لزيادة كفاءة أعمالها منها قطاعات التصنيع والصناعات الكيميائية، والتسويق، والفضاء، وغيرها حيث توجد الحاجة الى عملية تعدين بيانات تقليدية متطورة وتعمل بشكل فعّال لذا يجب أن تكون تقنيات تقنيب البيانات موثوقة وقابلة للتكرار من قبل أفراد الشركة ويتطلب وجود خبرة تقنيب البيانات، كذلك تقديم عملية قياسية عابرة للصناعة لتقنيب البيانات (CRISP-DM) لأول مرة في عام 1990م بعد أن خضعت لعدة ورش عمل، وساهمت أكثر من 300 منظمة. تعدين البيانات هي عملية لاكتشاف

البيانات الثمينة المخفية وذلك عن طريق تقييم كميات كبيرة من المعلومات المخزنة في مستودعات البيانات عن طريق استخدام تقنيات تعدين البيانات. (Jaiswal ,2024, P :1).



الشكل (4) تقنيات تنقيب البيانات (Jaiswal ,2024, P :1)

12-أهم أدوات تنقيب البيانات وأهميتها:

أهمية أدوات تنقيب البيانات مفتوحة تكمن في الوصول إلى مصادر بيانات مختلفة، حيث أن لديها القدرة على معالجة البيانات السابقة، ولديها القدرة على دمج تقنيات مختلفة، والعمل على مجموعات بيانات كبيرة، وتصور البيانات والنماذج بشكل جيد، ولها قابلية التوسيع، والتوافق مع أنظمة أخرى، ووجود مجتمع تطوير نشط.

KEEL: هي أداة تنقيب البيانات ولديها واجهة المستخدم الرسومية البسيطة لتدفق البيانات، ومتوفرة مجاناً

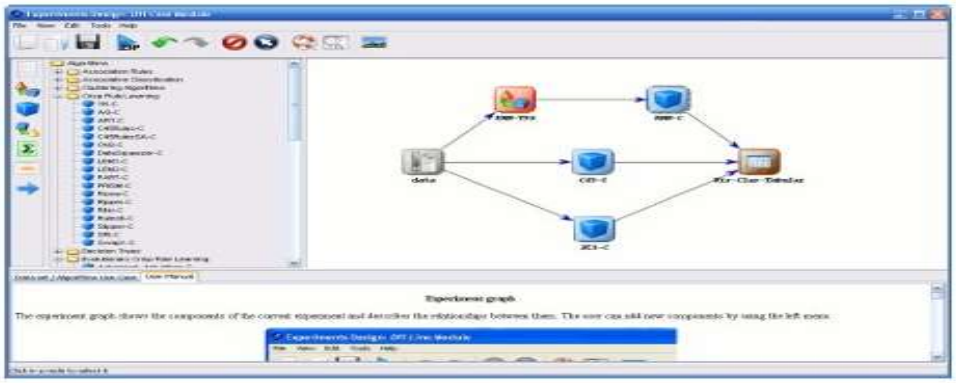


Figure 4: KEEL Data Mining Tool

الشكل 5 أداة KEEL

KNIME

تستخدم هذا الأداة لمعالجة البيانات قبل التحليل وتظهر أهميتها في تحليل بيانات العملاء والبيانات المالية والذكاء التجاري حيث أن لهذه الأداة القدرة لتحليل مجموعة بيانات تعليمية وتعمل على اتخاذ القرارات.

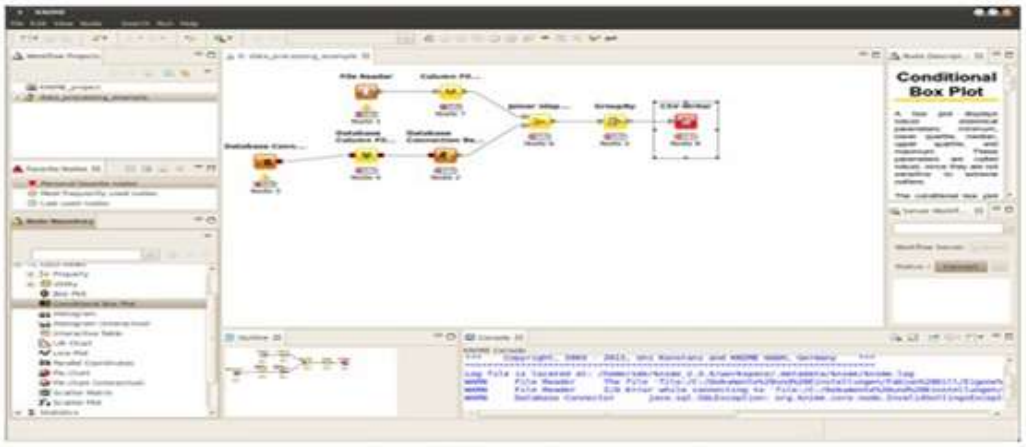


Figure 5: KNIME Data Mining Tool

الشكل 6 أداة KNIME

: هذه الأداة هي برنامج مفتوح المصدر المجاني **RapidMiner**

(FOOS) إصدار تجاري حيث تظهر أهميتها أن لديها ميزات تنقيب البيانات القياسية مثل التصفية والتجميع. يمكن دمج هذه الأداة مع لغات برمجة مثل لغة Python ولغة R.

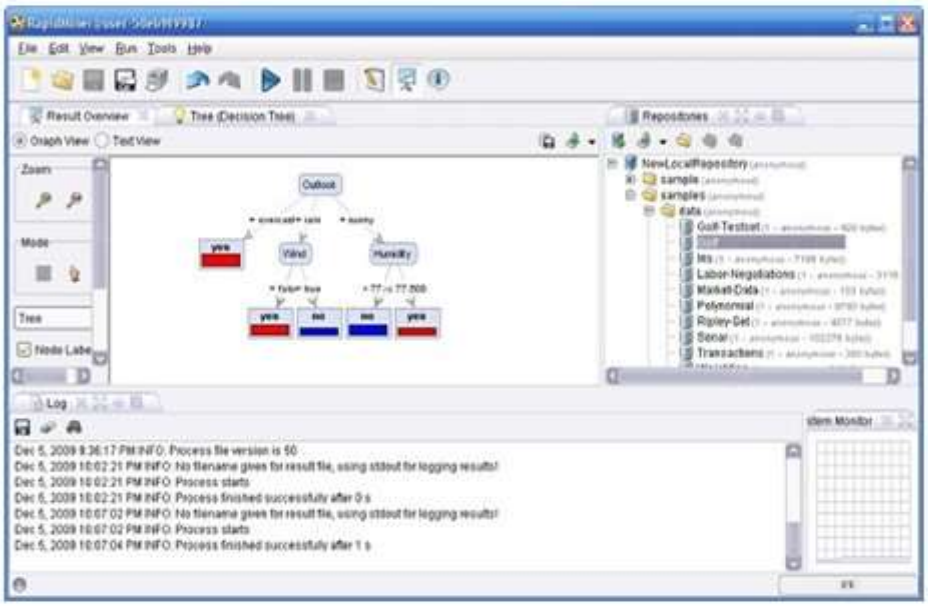


Figure 6: RapidMiner Data Mining Tool

الشكل 7 أداة RapidMiner

Weka: هي أداة مفتوحة المصدر المجانية معروفة يتم استخدامها لتقنيات تنقيب البيانات. تحتوي على مجموعة من خوارزميات التعلم الآلي، ويمكن تطبيق الخوارزميات المتاحة في هذه الأداة مباشرة على مجموعة البيانات المجهزة بتنسيق ARFF أو CSV أو من رمز

Javaيمكن الوصول إلى وظيفة Explorer من سطر الأوامر بمساعدةExperimenter ، يمكن للمستخدم مقارنة أداء مختلف خوارزميات تعلم الآلة على مجموعة من مجموعات البيانات.

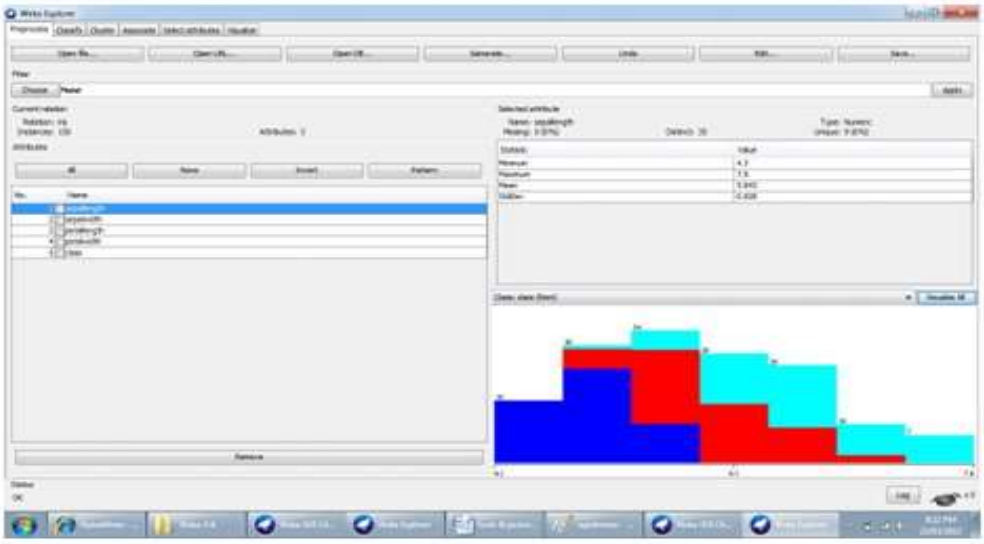


Figure 7: WEKA Data Mining Tool

الشكل8 أداة WEKA

Tanagra: هو أداة تنقيب في البيانات مفتوحة المصدر ومجانية لأغراض البحث الأكاديمية والتي تتضمن عدة أساليب لتحليل البيانات، ويستخدم في مجال التعلم الآلي وقواعد البيانات والغرض من هذه الأداة هو:

- يقوم بتوفير برمجيات تنقيب في البيانات للباحثين لتحليل أداء النماذج.
- يسمح للباحثين بدمج أساليب تنقيب في البيانات المعدلة.، كما يسمح لهم بمقارنة أدائهم.
- يتم نشر منهجية ممكنة لبناء هذا النوع من البرمجيات.

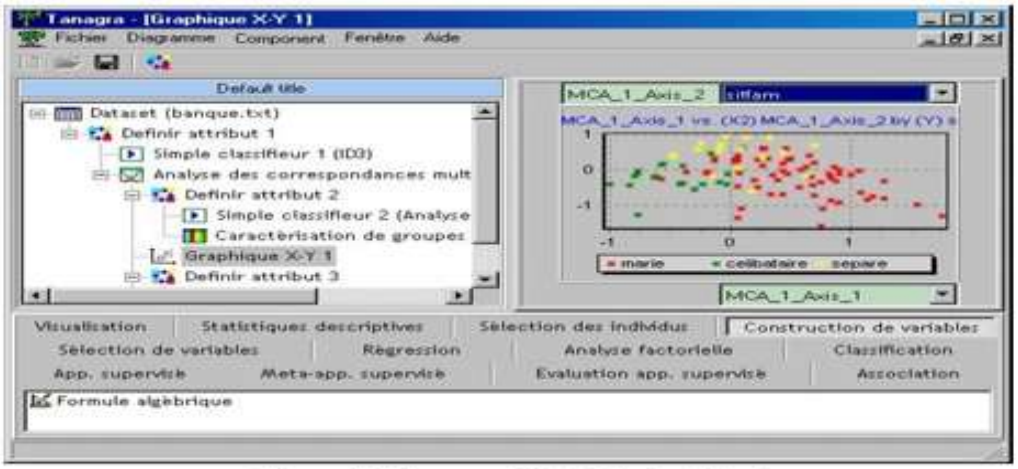


Figure 8: Tanagra Data Mining Tool

الشكل (9): أداة Teanagra

Orange

هي أداة مفتوحة المصدر لتنقيب البيانات تستخدم لتصور البيانات وتعلم الآلة وتحليل البيانات، تتكون Orange من لوحة يستطيع المستخدم من خلالها أن يضع فيها العناصر الرسومية حيث تشمل هذه العناصر الوظائف

الأساسية مثل قراءة مجموعة البيانات، واختيار الميزات المتاحة، وتدريب النماذج التنبؤية، وتصور عناصر البيانات، وغيرها (Dol &Jawandhiya ,2022,p: 381-385).

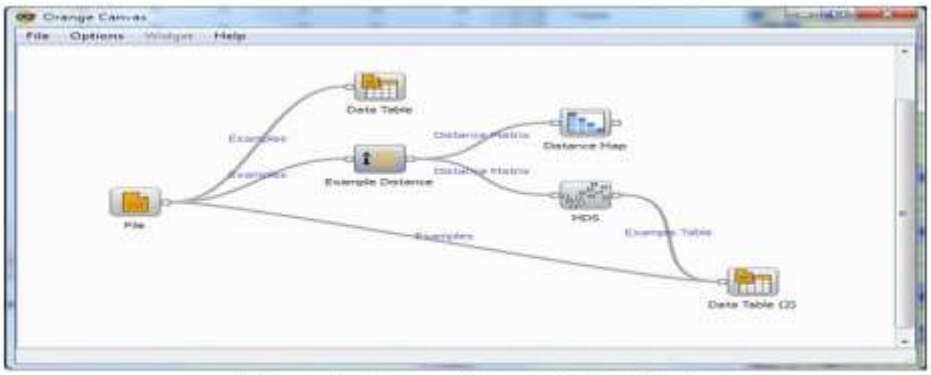


Figure 9: Orange Data Mining Tool

الشكل (10) أداة Orange

النظام	Orange	Tanagra	Weka	RapidMiner	KNIME	KEEL
المطور	University of Ljubljana	Lumière University Lyon 2	جامعة وايكاتو University of Waikato, 2002	RapidMiner, 2001	KNIME.com AG 2006	المشاريع الوطنية الأسبانية
لغة الكتابة	بايثون,Python, C++, C	C++	جافا	جافا	جافا	لغة البرمجة الإحصائية R statistical programming
النوع	مفتوح المصدر، مجاني	مفتوح المصدر مجاني	مفتوح المصدر	تجاري، مجاني مع خيارات دفع	مفتوح المصدر، مجاني	Data Mining
نظام التشغيل	لينكس، ماك أو إس، ويندوز Cross platform	لينكس، ماك أو إس، ويندوز	لينكس، ماك أو إس، ويندوز	منصات متعددة	لينكس، ماك أو إس، ويندوز	لينكس، ماك أو إس، ويندوز
الخوارزميات التي تدعم تنقيب البيانات	تستخدم لتحليل البيانات في التعليم الالكتروني يتم تطبيق مجموعة من الخوارزميات، التصنيف، التجميع وقواعد الارتباط على مجموعة من البيانات الدراسية لتحليل أداء الطلاب وتحسين العملية التعليمية وسهولة الاستخدام، تركيز التحليل الاستكشافي، مجتمع النشط يمكن للباحثين اختبار خوارزميات	تستخدم للتحليل الإحصائي وتعلم الآلة والتنقيب عن البيانات صعب على المبتدئين، قد يكون التحليل المتقدم صعبًا، قواعد الارتباط، خوارزميات البناء، التحليل متعدد التغيرات	تعلم الآلة، وتنقيب البيانات، والتحليل الإحصائي، والتصور التصنيف، الانضمام الاتحاد، التجميع الانحدار، قواعد الارتباط تحتوي Weka على خوارزميات لمعالجة البيانات السابقة، والتصنيف مثل Naïve Bays و Logistic Regression	علم البيانات تنظيف البيانات، تستخدم بسرعة، بيانات ضخمة، نشر سهل للتقارير، تحليل، النماذج الحديثة، تحليلات البيانات التنبؤية حيث أن هذه الأداة تستخدم في البحوث وتم تطبيق تقنية التصنيف ID3 مع مقياس الأداء Accuracy وخوارزمية الارتباط	إعداد تقارير متخصصة، ذكاء الأعمال، التنقيب عن البيانات العميقة، التعلم، تحليل البيانات، التنقيب عن النصوص، وتحليلها، جمعها والنمذجة، وتسجيل البيانات، سهولة الاستخدام، قابلة للتوسعة، مجموعة من الأدوات، واجهة سحب وإفلات، قد تكون واجهة	الإحصاءات، التجمع، النمذجة، التصور واجهة مستخدم سهلة الاستخدام يتم استخدام هذه الواجهة لتجربة مجموعات بيانات مختلفة التي تفحص سلوك خوارزميات مختلفة مثل التصنيف والتجميع والانحدار وتنقيب الأنماط.

	التعلم الآلي على مجموعة البيانات. على سبيل المثال في قطاع التعليم لتطبيق تقنية التصنيف باستخدام نماذج Naïve Bayes وRandom Forest وNeural Network		و ADTree وغيرها، والانحدار، والتجميع مثل Expectation Maximizatio n، K-means وقواعد الارتباط مثل Apriori وFPGrowth	وتم Apriori استخدام هذه الأداة لتقنيات التصنيف مثل Naïve Bayes و Deep Learning in Neural Network و Decision Tree	المستخدم معقولة، بحيث تكون أقل تكلفة تحليل، إعداد التقارير، ذكاء البيانات، توفر القابلة برامج للتنزيل مثل KNIME Analytic Platform وKNIME Server وKNIME Extension وKNIME Supported Extension وCommunity and Partner Extension.	
أدوات استخراج البيانات التجارية	Oracle Data mining	STATISTIC A	SAS Enterprise	Microsoft Analysis services	KXEN	SPSS
الترخيص	العمومي GNU (General Public License)	مجاني	عمومًا GNU GNU General Public License	AGPL (Community Edition) Closed (Enterprise Edition)	ترخيص عام	جنو الترخيص الإصدار 3

الموقع الإلكتروني	https://orangedatamining.com/	http://eric.univ lyon2.fr/~ricco/tanagra /en/tanagra.html	https://www.cs.waikato.ac.nz/ml/weka/courses.html	https://altair.com/altair-rapidminer	https://www.knime.com/software-overview https://knime.co/	http://www.keel.es/

الجدول (2) من إعداد الباحثة بالاعتماد على المصدر: (Dol &Jawandhiya ,2022,p: 381-385)

تنقيب الويب(Web Mining) هو تنقيب الويب هو فرع من تنقيب البيانات يركز على استخراج المعلومات المفيدة من بيانات الويب، ويصنف بشكل أساسي إلى ثلاثة أنواع: تنقيب محتوى الويب، تنقيب هيكل الويب، وتنقيب استخدام الويب.

أهمية تقنية تنقيب الويب:

يعتبر تصنيف الإعلانات عبر الإنترنت حسب المهارات والتصنيفات الوظيفية مجال مهم حيث استخدم بشكل واسع تحليل الإعلانات الوظيفية على الإنترنت، المعروفة اختصارًا بـOJAs ، التي تقدم الأساس الجيد لمطابقة الوظائف الشاغرة مع المتقدمين المناسبين، إن هذه الإعلانات تم نشرها في قواعد بيانات على الإنترنت مثل Monster أوStepstone ، وكذلك في قواعد بيانات الهيئات الرسمية مثل الوكالة الفيدرالية للتوظيف في ألمانيا ، تحتوي هذه الإعلانات على معلومات متنوعة مثل بيانات الشركة الموظفة والوظيفة المعلنة ومتطلبات العمل المطلوبة. أن التقنيات المستخدمة لاستخراج المهارات من الإعلانات تتراوح من التصنيف الآلي للمصطلحات إلى استخدام نماذج اللغة الكبيرة كما تم الاهتمام بشكل خاص بتطوير أطر تصنيف المهارات واستخدام تقنيات البيانات الكبيرة **Fischer, & Dörpinghaus ,2024, (P.52).**

الدراسات السابقة عن تنقيب البيانات وتنقيب الويب:

تزايد الاهتمام باستخراج البيانات من الويب، مثل، قواعد البيانات التعليمية والإعلانات وأنظمة المعلومات حيث يشير تعدين الويب إلى تطبيق تقنيات التنقيب عن البيانات لاكتشاف واستخلاص الأنماط والمعرفة من بيانات الويب حيث يدعم عملية صنع القرار وإدارة العمليات في التعليم وتتمثل التحديات العامة في الاستخراج الآلي للمعرفة من البيانات (المقاطع المفسرة من النصوص) ورسم الخرائط لمجموعات البيانات الموجودة، كما توجد العديد من التحديات المتعلقة بتكامل البيانات.

1-دراسة (Fischer, & Dörpinghaus ,2024):

ناقشت الدراسة كيف ساهم تصنيف الإعلانات عبر الإنترنت حسب المهارات والتصنيفات الوظيفية مجال مهم حيث استخدم بشكل واسع تحليل الإعلانات الوظيفية على الإنترنت، المعروفة اختصارًا بـOJAs ، التي تقدم الأساس الجيد لمطابقة الوظائف الشاغرة مع المتقدمين المناسبين، إن هذه الإعلانات تم نشرها في قواعد بيانات على الإنترنت مثل Monster أوStepstone ، وكذلك في قواعد بيانات الهيئات الرسمية مثل الوكالة الفيدرالية للتوظيف في ألمانيا ، تحتوي هذه الإعلانات على معلومات متنوعة مثل بيانات الشركة الموظفة والوظيفة المعلنة ومتطلبات العمل المطلوبة ، كما أوضحت الدراسة أن التقنيات المستخدمة لاستخراج المهارات من الإعلانات تتراوح من التصنيف الآلي للمصطلحات إلى استخدام نماذج اللغة الكبيرة كما تم الاهتمام بشكل خاص بتطوير أطر تصنيف المهارات واستخدام تقنيات البيانات الكبيرة.

وبحسب نموذج KSAO، الذي طوره فيشر ونيوبرت، اعتبروا أن المهارات والمعرفة والقدرات مكونات أساسية لكفاءة الأداء في العمل، وهي متميزة عن بعضها البعض ، كما وجدوا التحديات المتعلقة بكيفية استخدام البيانات المشتركة وتقنيات استخراج البيانات ، حيث تم الاعتماد على الخبرات السابقة مع نظم مثل BERUFENET و OJAsوالتصنيفات المختلفة للمهارات في النصوص الألمانية، ولكن واجه الباحثان نقص في الأبحاث حول كيفية ربط مصادر البيانات المتنوعة عبر الإنترنت، كما وجدوا عدة تحديات خاصة بالسوق الألماني التي تتطلب المزيد من البحث، مثل ربط لوائح التدريب بإعلانات التعليم والتدريب المهني.

استنتجت الدراسة أن تنقيب الويب مهم لاستخراج وتحليل البيانات من مصادر عديدة مثل الإعلانات عن الوظائف وبوابات البيانات الحكومية حيث تم استخدامه في تحليل السوق العمل، وتحديد متطلبات الوظائف والفرص التدريبية المتاحة، كما تم استخدام تقنيات متقدمة مثل معالجة اللغات الطبيعية وتقنيات الرسم البياني المعرفي لربط البيانات واستخراج العلاقات والكيانات، حيث أستخدم الباحثان الأطر التصنيفية مثل KldB و ESCOو ISCO لتنظيم وتصنيف المهن والمسارات التعليمية ، وكانت من أبرز التحديات الرئيسية هو التعامل مع بيانات غير مهيكلة ومتنوعة الصيغ، وتم التغلب على هذه التحديات باستخدام التصنيفات المتعددة والمطابقة بناءً على الأكواد مثل .KldB/DKZ .

تم استخدام رموز KldB/DKZ لربط بيانات الوظائف والتعليم والتدريب المهني عبر المصادر المختلفة، ويساعد هذا الربط في إنشاء شبكات معرفية تظهر مسارات التعليم والتطور المهني، تم استخدام البيانات المستخرجة لتحسين الفهم العلمي لسوق العمل وتطوير برامج تعليمية وتدريبية حسب الاحتياجات الفعلية، وتوسع استخدام بيانات تنقيب الويب لتشمل تحليلات أعمق للمتطلبات المهنية وتحديث المسارات التعليمية والتدريبية حسب تغيرات سوق العمل.**(Fischer, at al.,2024)**.

2-دراسة (Bhavana at al.,2022):

ناقشت الدراسة نظام توصية مختلط معروف بـ AERS (نظام التوصية المرن والفعال القائم على تنقيب الويب)، الذي يجمع بين تقنيات تصفية المحتوى والتصفية التعاونية وذلك لتحسين التوصيات، حيث ناقشت الدراسة المفاهيم الرئيسية المتعلقة بتنقيب البيانات وتنقيب الويب:

1-تصفية المحتوى:(Content Filtering)

يعتمد هذا النوع من الفلترة على وصف العناصر، مثل الغرض من العنصر، المؤلف، الفنان، والكلمات الرئيسية، ويتم بناء ملف تعريف لكل مستخدم بحسب تفضيلاته والتفاعلات السابقة مع النظام، ويستخدم هذا الملف لتوقع العناصر التي قد يفضلها المستخدم.

2-التصفية التعاونية:(Collaborative Filtering)

تنتج التوصيات حسب سلوك المستخدم السابق، مثل العناصر التي شاهدها أو قيّمها، وكذلك القرارات المماثلة التي اتخذها مستخدمون آخرون بالإضافة أنه يتم تقسيم هذا النوع إلى التصفية التعاونية القائمة على العنصر والقائمة على المستخدم، حيث يتم العثور على التشابه بين المستخدمين والعناصر ثم تقديم التوقعات بالاستناد على هذه التشابهات.

3- تنقيب الويب(Web Usage Mining) يشير إلى استخراج المعرفة من الأنشطة التنقلية للمستخدمين على موقع ويب معين وأيضا يستفيد من ملفات السجل لتحليل أنماط الوصول إلى المستخدمين وتحسين تجربة المستخدم. ويستخدم في النظام المختلط AERS للتنبؤ بالعناصر التي قد يهتم بها المستخدمون حسب على ملفاتهم الشخصية.

4- النموذج الهجين :AERS: Adaptive and Recommendation System Based on Web Usage Mining Efficient Hybrid ، يجمع بين التصفية التعاونية وتصفية المحتوى لتعزيز دقة التوصيات، ويستخدم تقنيات مختلطة مثل النهج المختلط، النظام الموزون، التقنية القائمة على التبديل، ودمج الميزات لتحسين الأداء. أظهرت نتائج الدراسة أن النظام المقترح AERS يعمل بأداء أفضل إذا تم مقارنته بالنظم الفردية من حيث الدقة والتنوع في التوصيات، مما يدل على فعالية النموذج الهجين في توفير توصيات مخصصة ودقيقة للمستخدمين (Bhavana at al.,2022, P.670-671).

3-أشارت دراسة2022 Liu& Liu الى ضرورة استخدام تنقيب الويب والبيانات في إطار تعليم عن بعد بهدف تصميم الدورات وإعادة تنظيم المواقع التعليمية لتصبح بيئة التعليم فعالة مناسبة للمستخدمين، فيتحقق التعليم الذي يتناسب مع قدرات واحتياجات الطلاب المختلفة.

تنقيب في الويب هو تحليل محتوى المستندات واستخدام الموارد المتاحة، والعلاقات بين الموارد لاكتشاف أنماط وقواعد فعالة وجديدة ومن المحتمل أن تكون ذات قيمة ومفهومة، هناك نوعان من تقنيات التعدين الرئيسية التي يتم بحثها وهي: تعدين استخدام الويب وتعدين محتوى الويب.

تعدين استخدام الويب (استخراج معلومات الوصول إلى الويب) هو استخراج سجلات الوصول التي تركها الخادم عندما يصل المستخدم إلى الويب، أي استخراج وضع الوصول للمستخدم الذي يصل إلى موقع الويب للعثور على نمط التصفح للمستخدم الذي يصل إلى الموقع وتكرار الوصول إلى الصفحة والمعلومات الأخرى؛ الكائنات التي سيتم استخراجها هي سجلات على الخادم بما في ذلك بيانات سجل الخادم وما إلى ذلك وتشمل أساليب التعدين تحليل المسار؛ اكتشاف قواعد الارتباط وأنماط التسلسل؛ التجميع والتصنيف. تعدين محتوى الويب (Web Content Mining)هو التنقيب عن محتوى صفحة الويب. يمكن تقسيم عملية التنقيب عن المحتوى إلى فئتين: التنقيب عن المستندات النصية بما في ذلك TEXT و HTMLوغيرها من التنسيقات والتنقيب عن مستندات الوسائط المتعددة بما في ذلك أنواع الوسائط مثل الصور والصوت والفيديو وما إلى ذلك. يمكن للتنقيب عن نص الويب تلخيص محتوى عدد كبير من مجموعات المستندات على الويب وتصنيفه وتجميعه وربطه وتحليله، واستخدام مستندات الويب للتنبؤ بالاتجاه كما تشير تصنيف المستندات إلى تحديد فئة لكل وثيقة في مجموعة المستندات

حسب لفئة موضوع محددة في السابق وبهذه الطريقة، لا يمكن للمستخدمين تصفح المستندات بسهولة فحسب، بل يستطيعوا تسهيل العثور على المستندات عن طريق تقييد نطاق البحث ، إن الفرق بين تجميع المستندات وتصنيفها هو أن التجميع لا يحتوي على فئة موضوع محددة مسبقًا ، الهدف منها هو تقسيم مجموعة المستندات إلى عدة مجموعات، ويشترط أن يكون تشابه محتوى المستند في نفس المجموعة أكبر قدر ممكن التشابه بين المجموعات المختلفة صغير قدر الإمكان.، يشير تحليل الارتباط إلى اكتشاف العلاقة بين الكلمات المختلفة من مجموعة من الوثائق ، إن تقنيات التنقيب عن الويب تشمل:1-تنقيب استخدام الويب(Web Usage Mining) يفيد في تعدين سجلات الوصول التي يتركها الخادم عندما يتصفح المستخدمون الويب، بهدف اكتشاف نمط تصفح المستخدم للموقع وتردد زيارة الصفحات وتشمل الأدوات المستخدمة له تحليل المسارات، اكتشاف قواعد الارتباط والأنماط التسلسلية، التجميع، والتصنيف.

2-تنقيب محتوى الويب(Web Content Mining) يتعلق بتعدين محتوى صفحات الويب وينقسم إلى تعدين الوثائق النصية وتعدين الوثائق متعددة الوسائط ويتم تلخيص النصوص، تصنيفها، تجميعها، وربطها وتحليلها لاستخدام الوثائق الويب في التنبؤ بالاتجاهات حيث يشمل التصنيف تحديد فئة لكل وثيقة حسب فئة موضوع محددة مسبقًا، مما يسهل على المستخدمين تصفح الوثائق والعثور عليها ، حيث أن التجميع لا يعتمد على فئات موضوعية محددة في السابق ويهدف إلى تقسيم مجموعة الوثائق إلى مجموعات بحيث تكون الوثائق داخل كل مجموعة متشابهة ، بالإضافة أن تحليل الارتباط يشمل اكتشاف العلاقات بين الكلمات المختلفة من مجموعة وثائق.

هذه التقنيات مفيدة وتستخدم في مختلف جوانب الويب مثل التجارة الإلكترونية، الردود الآلية للبريد الإلكتروني، تصميم المواقع الشخصية، والتقارير الويب ويعتبر تنقيب استخدام الويب محور أساسي من الذكاء التجاري في الشركات ويعتمدوا عليه في اتخاذ القرارات التجارية من خلال الاستخدام المتميز لخدمات الويب، وهو ضروري لإدارة علاقات العملاء لتوفيره بيانات حول سلوك العملاء حتى نقطة التفاعل بين الشركة والعميل(Liu & Liu 2022) .

4-ناقشت الدراسة(Guo,2021) تنقيب الويب بشكل شامل، حيث تم استخدام أفكار وأساليب تنقيب البيانات التقليدية لاستخراج الأنماط والمعلومات المفيدة والمحتملة من مصادر الويب وأنشطته. يتم تطبيق تنقيب الويب في إدارة المعلومات ودعم القرار والتحكم في العمليات وكذلك صيانة البيانات. أشارت الدراسة الى الدور المحوري للزواحف الويبية في كامل النظام، حيث تعتبر مصدر البيانات الذي يحدد ما إذا كان محتوى النظام غني ومحدث بالمعلومات في الوقت المناسب، وذكرت الدراسة أنواع التنقيب عن الويب:

1-تنقيب استخدام الويب(Web Usage Mining) يتم تعدين سجلات الوصول التي يتركها المخدم عندما يتصفح المستخدمون الويب لاستخراج نمط تصفح المستخدم وتردد الصفحات، ويشمل ذلك تحليل المسار، اكتشاف قواعد الارتباط والأنماط التسلسلية، التجميع والتصنيف.

2-تنقيب محتوى الويب(Web Content Mining) يتضمن تعدين محتوى صفحات الويب وينقسم إلى فئتين: تعدين الوثائق النصية، النصوص، HTML وتعدين الوثائق متعددة الوسائط (الصور، الصوت، الفيديو وغيرها).

-التصنيف الوثائقي: يتم تحديد فئة لكل وثيقة في مجموعة الوثائق وفقاً لفئة موضوعية محددة مسبقاً، مما يسهل على المستخدمين التصفح والعثور على الوثائق.

-التجميع الوثائقي: يختلف عن التصنيف بعدم وجود فئة موضوعية محددة مسبقاً، حيث يهدف إلى تقسيم مجموعة الوثائق إلى مجموعات، بحيث تكون الوثائق في كل مجموعة متشابهة قدر الإمكان ومختلفة عن الوثائق في المجموعات الأخرى.

-تحليل الارتباط: يهدف إلى اكتشاف العلاقات بين الكلمات المختلفة من مجموعة من الوثائق.

هذه الأنواع من التنقيب تعمل على تحسين جودة بحث المعلومات وتنقيب البيانات من خلال تزويد المصممين بالمعلومات التي تساعد على تحسين تصميم المناهج وإعادة تنظيم المواقع الإلكترونية بطريقة تجعل التعليم أكثر فعالية وملاءمة للمستخدمين.

ناقشت الدراسة طريقة تنقيب البيانات عن المعلومات الويبية باستخدام الزاحف الموضوعي. حيث تقدم هيكلية لبحث المعلومات الويبية وتنقيب البيانات، وتعرض التقنية الأساسية ومبدأ العمل لهذه الهيكلية. بعد تحليل وظائف وعيوب الزاحف العادي، ركزت الدراسة على مبدأ عمل الزاحف الموضوعي، طريقة التنفيذ، وتحليل الأداء، بالإضافة إلى الوظائف التي تميز هذا الزاحف عن الزواحف الأخرى وتطبيقه في نظام بحث وتنقيب المعلومات الويبية. النتائج التجريبية تظهر أن الزاحف قادر على جمع كافة أنواع الموارد المعلوماتية على شبكة الإنترنت، وهو ما يفيد في مراقبة وإدارة المحتوى الثقافي للشبكة.

الكلمات المفتاحية: شبكة موضوعية، زاحف، تنقيب البيانات، بحث معلومات الويب.

5- دراسة (Malviya, 2 015):

تنقيب البيانات(Data Mining) هو عملية استخراج بيانات قيمة من قواعد بيانات ضخمة حيث يتم استخدام تقنيات تنقيب البيانات لفهم الأنماط والاتجاهات والعلاقات داخل مجموعات بيانات كبيرة للمساعدة في اتخاذ قرارات، إن الويب، كمصدر غني بالبيانات، يوفر فرصة كبيرة لاستخدام تنقيب البيانات لاستكشاف سلوكيات المستخدمين وتحسين الخدمات والمنتجات.

تنقيب الويب(Web Mining) هو تنقيب الويب هو فرع من تنقيب البيانات يركز على استخراج المعلومات المفيدة من بيانات الويب، ويصنف بشكل أساسي إلى ثلاثة أنواع: تنقيب محتوى الويب، تنقيب هيكل الويب، وتنقيب استخدام الويب.

- تنقيب استخدام الويب(Web Usage Mining) هو تنقيب الويب الذي يعتمد على استخراج بيانات من سجلات الويب لفهم وتحسين تجربة المستخدم ويتضمن ثلاث مراحل رئيسية:

ما قبل المعالجة(Preprocessing) تنظيف البيانات وتحديد الجلسات، والتي تعد خطوة أساسية لتحضير البيانات للتحليل.

كشف الأنماط(Pattern Discovery) استخدام تقنيات تنقيب البيانات المتقدمة لتحديد الأنماط والاتجاهات في سلوك المستخدم.

تعلم الأنماط(Pattern Learning) تحليل الأنماط المكتشفة لفهم السلوكيات وتطبيق النتائج في تحسين تجربة المستخدم.

تنقيب استخدام الويب له تطبيقات عديدة مثل تخصيص المواقع الإلكترونية، تحسين أداء النظام، إعادة هيكلة الموقع، ودعم التصميم الأمثل للمواقع الإلكترونية. يتم استخدام نتائج تنقيب استخدام الويب لتوفير تجربة مستخدم محسنة ولتقديم توصيات بحسب الأنماط المكتشفة في سلوك المستخدم، تظهر أهمية التعدين في استخدام الويب في مجال التجارة الإلكترونية وإدارة العلاقات مع العملاء(CRM) وتم تلخيصها في نقاط الرئيسية التالية:

1- المواقع الإلكترونية التكيفية تمثل خطوة إضافية في تطوير الويب، حيث يمكن إعادة تنظيم المحتوى وهيكل الموقع استجابةً لسلوك المستخدمين الذي يتم تحليله عن طريق تقنيات التعدين في استخدام الويب.

2- التجارة الإلكترونية عبر التعدين في استخدام الويب حيث يعتبر التعدين في استخدام الويب أداة مهمة.

استنتجت الدراسة أن نموذج التعدين في استخدام هو تطبيق لأساليب التنقيب عن البيانات في مناطق تخزين ضخمة لسجلات الويب لتحديد المعرفة القيمة المتعلقة بالأنماط السلوكية للعميل ومعلومات استخدام موقع الويب لتسهيل مجموعة متنوعة من مجالات تصميم مواقع الويب يستفيد من سجلات الخوادم، حيث أن التعدين له دور مهم في استخدام الويب وفي فهم قابلية استخدام تصميم الموقع الإلكتروني، تحسين علاقات العملاء، وتعزيز أداء النظام بشكل عام، كما يوفر التعدين في استخدام الويب الدعم لتخصيص الخدمات، تصميم المواقع، واتخاذ قرارات الأعمال الأخرى كما ناقشت الدراسة أيضا بعض المشكلات المرتبطة بالتعدين في استخدام الويب، مثل إيجاد المعلومات المرغوبة، والعثور على المعلومات المتعلقة، وتعلم المعرفة القيمة، والتوصيات أو التخصيص الشخصي للبيانات، والأعمال الحديثة في مجال أبحاث التعدين في استخدام الويب .

تبدأ بيانات سجل المستخدمين بشكل جماعي بموارد مختلفة كخوادم بديلة وجانب العميل وجانب الخادم وما إلى ذلك. المعالجة المسبقة للبيانات حيث يتم إجراء سلسلة من معالجة ملف سجل الويب الذي يغطي تنظيف السجلات والتعرف على الجلسة والتعرف على العملاء وموضع إنهاء المسار واكتشاف المعاملات.

أنواع تقنيات التنقيب في الويب:

تنقيب في الويب هو تحليل محتوى المستندات، واستخدام الموارد المتاحة، والعلاقات بين الموارد لاكتشاف أنماط وقواعد فعالة وجديدة ومن المحتمل أن تكون ذات قيمة ومفهومة، هناك نوعان من تقنيات التعدين الرئيسية التي يتم بحثها وهي: تعدين استخدام الويب وتعدين محتوى الويب.

تعدين استخدام الويب (استخراج معلومات الوصول إلى الويب) هو استخراج سجلات الوصول التي تركها الخادم عندما يصل المستخدم إلى الويب، أي استخراج وضع الوصول للمستخدم الذي يصل إلى موقع الويب للعثور على نمط التصفح للمستخدم الذي يصل إلى الموقع وتكرار الوصول إلى الصفحة والمعلومات الأخرى؛ الكائنات التي سيتم استخراجها هي سجلات على الخادم بما في ذلك بيانات سجل الخادم وما إلى ذلك، وتشمل أساليب التعدين تحليل المسار؛ اكتشاف قواعد الارتباط وأنماط التسلسل؛ التجميع والتصنيف ،تعدين محتوى الويب (Web Content Mining)هو التنقيب عن محتوى صفحة الويب يمكن تقسيم عملية التنقيب عن المحتوى إلى فئتين: التنقيب عن المستندات النصية بما في ذلك TEXT و HTMLوغيرها من التنسيقات والتنقيب عن مستندات الوسائط المتعددة بما في ذلك أنواع الوسائط مثل الصور والصوت والفيديو وما إلى ذلك. يمكن للتنقيب عن نص الويب تلخيص محتوى عدد كبير من مجموعات المستندات على الويب وتصنيفه وتجميعه وربطه وتحليله، واستخدام مستندات الويب للتنبؤ بالاتجاه كما تشير تصنيف المستندات إلى تحديد فئة لكل وثيقة في مجموعة المستندات

حسب لفئة موضوع محددة في السابق وبهذه الطريقة، لا يمكن للمستخدمين تصفح المستندات بسهولة فحسب، بل يستطيعوا تسهيل العثور على المستندات عن طريق تقييد نطاق البحث ، إن الفرق بين تجميع المستندات وتصنيفها هو أن التجميع لا يحتوي على فئة موضوع محددة مسبقًا ، الهدف منها هو تقسيم مجموعة المستندات إلى عدة مجموعات، ويشترط أن يكون تشابه محتوى المستند في نفس المجموعة أكبر قدر ممكن التشابه بين المجموعات المختلفة صغير قدر الإمكان.، يشير تحليل الارتباط إلى اكتشاف العلاقة بين الكلمات المختلفة من مجموعة من الوثائق . (Liu & Liu 2022, P.488)

- تنقيب استخدام الويب(Web Usage Mining) : هو تنقيب الويب الذي يعتمد على استخراج بيانات من سجلات الويب لفهم وتحسين تجربة المستخدم ويتضمن ثلاث مراحل رئيسية:

ما قبل المعالجة(Preprocessing) تنظيف البيانات وتحديد الجلسات، والتي تعد خطوة أساسية لتحضير البيانات للتحليل.

كشف الأنماط(Pattern Discovery) استخدام تقنيات تنقيب البيانات المتقدمة لتحديد الأنماط والاتجاهات في سلوك المستخدم.

تعلم الأنماط(Pattern Learning) تحليل الأنماط المكتشفة لفهم السلوكيات وتطبيق النتائج في تحسين تجربة المستخدم.

تنقيب استخدام الويب له تطبيقات عديدة مثل تخصيص المواقع الإلكترونية، تحسين أداء النظام، إعادة هيكلة الموقع، ودعم التصميم الأمثل للمواقع الإلكترونية. يتم استخدام نتائج تنقيب استخدام الويب لتوفير تجربة مستخدم محسنة ولتقديم توصيات بحسب الأنماط المكتشفة في سلوك المستخدم، تظهر أهمية التعدين في استخدام الويب في مجال التجارة الإلكترونية وإدارة العلاقات مع العملاء(CRM) وتم تلخيصها في نقاط الرئيسية التالية:

1- المواقع الإلكترونية التكيفية تمثل خطوة إضافية في تطوير الويب، حيث يمكن إعادة تنظيم المحتوى وهيكل الموقع استجابةً لسلوك المستخدمين الذي يتم تحليله عن طريق تقنيات التعدين في استخدام الويب.

2-التجارة الإلكترونية عبر التعدين في استخدام الويب حيث يعتبر التعدين في استخدام الويب أداة

.(Malviya& Agrawal, 2015 ,P.935-938)

<u>مراحل التنقيب عن استخدام الويب:</u>

من أجل حل بعض المشاكل الموجودة في تكنولوجيا المعالجة المسبقة للبيانات التقليدية لاستخراج سجلات الويب، تم استخدم تقنية محسنة لمعالجة البيانات المسبقة حيث يتم اعتماد استراتيجية تحديد الهوية حسب صفحة الويب في مرحلة تحديد هوية المستخدم، وهي أكثر فعالية من التقليدية ، في مرحلة تحديد الجلسة، يتم البدء في الإستراتيجية القائمة على عتبة مسبقة ثابتة مع إعادة بناء الجلسة أولاً، يتم تطوير مجموعة الجلسة الأصلية باستخدام تقنية العتبة المسبقة الثابتة، ثم يتم تحسين مجموعة الجلسة الأصلية باستخدام إعادة بناء الجلسة، أثبتت التجارب أن المعرفة المتقدمة بالمعالجة المسبقة للبيانات يمكن أن تحسن جودة نتائج المعالجة المسبقة للبيانات.

في تعدين استخدام الويب، يكون دور تجميع جلسات الويب مهم لتصنيف زوار الويب على أساس سجل نقرات المستخدم وتقييم التشابه وتساعد مجموعات جلسات الويب المستندة إلى Swarm بطرق مختلفة في إدارة موارد الويب بنجاح مثل تخصيص الويب وتعديل المخطط وتغيير موقع الويب وتقديم خادم الويب (Malviya& Agrawal, 2015 ,P.935-938).

وفيما يلي يوضح الجدول بعض الفرق بين التنقيب في البيانات والتنقيب في الويب:

	(Web Mining) التنقيب عن الويب	**(Data Mining) التنقيب عن البيانات**
التعريف	**استخدام تقنيات التنقيب عن البيانات وأساليب التحليل لاستكشاف البيانات المستخرجة من صفحات الويب وخدمات الإنترنت 2015 (Malviya& Agrawal, P.935-938)**	**عملية استكشاف وتحليل كميات كبيرة من البيانات لاكتشاف أنماط مفيدة ومعلومات جديدة (Malviya& Agrawal, 2015 ,P.935-938)**

الهدف	فهم سلوك المستخدمين، وتحليل المحتوى والبنية لصفحات الويب لتحسين خدمات الويب.	اكتشاف الأنماط والعلاقات غير المتوقعة بين مجموعات البيانات. -الحصول على البيانات قائمة على المعرفة حيث يساعدها على اتخاذ القرارات. -التنقيب عن البيانات من حيث التكلفة يعتبر أفضل من غيرها من التطبيقات الإحصائية. إن عملية التنقيب عن البيانات تسهل اكتشاف للأنماط المخفية وتسهل على المستخدمين الجدد. -تساعد تنقيب البيانات في الكشف عن الاحتيال ومراقبة الأداء التشغيلي .(Khullar, 2023, p.2)
مصدر البيانات	البيانات مستخرجة بشكل أساسي من الإنترنت وتشمل سجلات الويب، محتويات الصفحات، روابط الصفحات، وبيانات المستخدمين. صفحات الويب، ومدونات الويب، ووسائل التواصل الاجتماعي، ومصادر البيانات الأخرى ذات الصلة بالويب (Utkarsh,2023)	يمكن أن تكون البيانات من أي نوع مثل السجلات المالية، نتائج الأبحاث، البيانات الطبية، قواعد البيانات ومستودعات البيانات ومستودعات البيانات الأخرى (Utkarsh,2023)
الأهمية	يستخدم في تحليل السوق العمل، وتحديد متطلبات الوظائف والفرص التدريبية المتاحة. التقنيات المستخدمة لاستخراج المهارات من الإعلانات تتراوح من التصنيف الآلي للمصطلحات إلى استخدام نماذج اللغة الكبيرة كما تم الاهتمام بشكل خاص بتطوير أطر تصنيف المهارات واستخدام تقنيات البيانات الكبيرة ، أن تنقيب الويب مهم لاستخراج وتحليل البيانات من مصادر عديدة مثل الإعلانات عن الوظائف وبوابات البيانات الحكومية حيث تم استخدامه في تحليل السوق العمل، وتحديد متطلبات الوظائف والفرص التدريبية المتاحة، كما تم استخدام تقنيات متقدمة مثل معالجة اللغات الطبيعية وتقنيات الرسم البياني المعرفي لربط البيانات واستخراج	تعتبر تقنية تنقيب البيانات من العمليات الحاسمة لاستراتيجيات التحليل الناجحة في المؤسسات، يستطيع أخصائي البيانات استخدام المعلومات التي تولدها في تطبيقات الذكاء التجاري والتحليلات المتقدمة التي تفحص البيانات المتدفقة أثناء إنشائها أو جمعها ويساعد تنقيب البيانات الفعال في مختلف جوانب تخطيط استراتيجيات الأعمال وإدارة العمليات، الوظائف التي تواجه العملاء مثل التسويق والإعلان والمبيعات ودعم العملاء، بالإضافة إلى التصنيع وإدارة سلسلة التوريد(SCM) ، والمالية والموارد البشرية .(HR) يدعم تنقيب البيانات اكتشاف الاحتيال، وإدارة المخاطر، وتخطيط الأمان السيبراني والعديد من حالات الاستخدام الأساسية الأخرى في الأعمال. وله دور هام في مجالات أخرى مثل الرعاية الصحية، والحكومة، والبحث العلمي،

	العلاقات والكيانات، حيث أستخدم الباحثان الأطر التصنيفية مثل KldB و ESCOو ISCO لتنظيم وتصنيف المهن والمسارات التعليمية Fischer, & Dörpinghaus (,2024, P.52).	والرياضيات والرياضة . (Jinyang,2023, P.10-11)
خصائص البيانات	بيانات غير منظمة في الغالب	البيانات المنظمة وشبه المنظمة وغير المنظمة
التطبيقات	في تحسين محركات البحث، التسويق الإلكتروني، تحليل تجربة المستخدم، وإدارة السمعة على الإنترنت.	يعتمد علماء البيانات من المحترفين وخبراء المجال في مجال الذكاء الاصطناعي والتحليلات على تنقيب البيانات، ويستطيع المحللين الأعمال الذين يتمتعون بالخبرة في مجال البيانات، والتنفيذيين، والعاملين الذين يعملون كعلماء بيانات في المؤسسة أيضًا أداء تنقيب البيانات، تشمل العناصر الأساسية لتنقيب البيانات التعلم الآلي والتحليل الإحصائي، بالإضافة إلى المهام الإدارية التي يتم تنفيذها لتحضير البيانات للتحليل-(Jinyang,2023, P.10 . 11) في الصناعات المختلفة مثل البنوك، التسويق، البحوث الطبية، التصنيع.
التحديات	عدم تجانس البيانات والغموض والطبيعة الديناميكية للويب.	التحديات جودة البيانات وقابلية التوسع ومخاوف الخصوصية، البيانات غير المتجانسة، المنهجية التي تستخدمها لاستخراج البيانات وجمعها، أمن البيانات، أخلاقيات جمع البيانات، خصوصية البيانات، البيانات المتفرقة، وقت الاستجابة وهو عبارة عن مسألة زمن استجابة نموذج التنبؤ حيث تعتمد فيها القرارات مثل تداول الأسهم (Martin ,2022, P :1- 4)
التقنيات	التنقيب عن النصوص، ومعالجة اللغات الطبيعية، وتحليل الصور، وتحليل الروابط	التجميع، التصنيف، قواعد الارتباط، الانحدار، شجرة القرار، الشبكات العصبية، القيم الشاذة، الأنماط المتسلسلة Akulwar.et al,2018, P:517-(518) الشبكات العصبية الصناعية، الخوارزمية الجينية، شجرة القرار ، قواعد الارتباط

		(Jinyang,2023, P.10-11)
	أدوات مثل Google Analyticsأدوات تحليل الروابط، برامج تعدين النصوص والمحتوى.	برامج متخصصة مثل SAS, RapidMiner, IBM SPSS Modeler.، ومصادر مفتوحة KEEL، KNIME، RapidMiner، Weka، Orange ,Tanagra Dol &Jawandhiya ,2022, p: (381-385)
العمليات	عملية التعدين على شبكة الإنترنت جمع البيانات- يتم جمع بيانات الويب من مصادر مختلفة، بما في ذلك صفحات الويب وقواعد البيانات وواجهات برمجة التطبيقات. المعالجة المسبقة للبيانات- تتم معالجة البيانات المجمعة مسبقًا لإزالة المعلومات غير ذات الصلة، مثل الإعلانات والمحتوى المكرر. تكامل البيانات- يتم دمج البيانات المعالجة مسبقًا وتحويلها إلى تنسيق منظم للتحليل اكتشاف النمط- يتم تطبيق تقنيات التنقيب على الويب لتحديد الأنماط والاتجاهات والعلاقات تقييم- يتم تقييم الأنماط المكتشفة لتحديد أهميتها وفائدتها التصور- يتم تصور نتائج التحليل من خلال الرسوم البيانية والمخططات والمرئيات الأخرى	تعريف المشكلات: تتضمن هذه المرحلة الأولية تحديد أهداف مشروع استخراج البيانات بوضوح. ويتضمن ذلك وضع معايير لقياس نتائج جهود استخراج البيانات. إنشاء قواعد بيانات التنقيب عن البيانات: يتم التركيز في هذه المرحلة على جمع البيانات وصيانتها وإعداد قواعد البيانات التي سيتم استخدامها في التنقيب. ويتضمن ذلك جمع البيانات ذات الصلة وضمان جودتها وإمكانية الوصول إليها. تحليل البيانات: يتضمن تحليل البيانات تحديد أهم مجالات البيانات التي ستؤثر على حل المشكلة. يتم اتخاذ القرارات بشأن البيانات التي يجب الاحتفاظ بها والبيانات التي يجب استبعادها هنا. إعداد بيانات النمذجة: تصفية واختيار المتغيرات والسجلات من قاعدة البيانات. ويتضمن أيضًا إنشاء متغيرات وحقول جديدة عن طريق تحويل البيانات الموجودة. وهذا أمر بالغ الأهمية لأنه يقوم بإعداد البيانات لوضع نماذج فعالة. النمذجة: اختيار خوارزميات استخراج البيانات المناسبة وتطبيقها لمعالجة البيانات. تتضمن هذه المرحلة فحص نماذج بيانات مختلفة لاختيار طريقة التنقيب عن البيانات الأكثر فعالية ودقة. تقييم وتفسير النموذج: بمجرد بناء النموذج، يجب تقييمه وتفسير نتائجه وقيمه.

	(Utkarsh,2023)	هذه الخطوة ضرورية لفهم فعالية النموذج وتحسينه إذا لزم الأمر. التنفيذ: بعد إنشاء النموذج والتحقق من صحته، يمكن أن تتبع مرحلة التنفيذ مسارين: الأول يتضمن تقديم خطة عمل وتوصيات بناءً على النموذج إلى المحللين لمزيد من التحليل والمرجع؛ أما الآخر فيتضمن المراقبة المستمرة لفعالية النموذج بعد التنفيذ للتأكد من بقائه فعالاً مع مرور الوقت-Yu &Li ,2021, P.330) . (331
الأنواع	أنواع التنقيب في الويب: تصنيف التعدين عبر الويب إلى 3 فئات فرعية: ➢ التعدين محتوى الويب ➢ التعدين استخدام الويب ➢ التعدين هيكل الويب (Shukla, et al.,2020,P.2) 1. تنقيب استخدام الويب(Web Usage Mining) ➢ يفيد في تعدين سجلات الوصول التي يتركها الخادم عندما يتصفح المستخدمون الويب، بهدف اكتشاف نمط تصفح المستخدم للموقع وتردد زيارة الصفحات. ➢ تشمل الأدوات المستخدمة له تحليل المسارات، اكتشاف قواعد الارتباط والأنماط التسلسلية، التجميع، والتصنيف. 2. 2-تنقيب محتوى الويب(Web Content Mining) ➢ يتعلق بتعدين محتوى صفحات الويب وينقسم إلى تعدين الوثائق النصية وتعدين الوثائق متعددة الوسائط. تلخيص النصوص، تصنيفها، تجميعها، وربطها وتحليلها لاستخدام الوثائق الويب في التنبؤ بالاتجاهات حيث يشمل التصنيف تحديد فئة لكل وثيقة حسب فئة موضوع محددة مسبقًا، مما يسهل على المستخدمين تصفح الوثائق والعثور عليها. ➢ التجميع: لا يعتمد على فئات موضوعية محددة في السابق ويهدف إلى تقسيم مجموعة الوثائق إلى مجموعات بحيث تكون الوثائق داخل كل مجموعة متشابهة. ➢ تحليل الارتباط: يشمل اكتشاف العلاقات بين الكلمات المختلفة من مجموعة وثائق. ➢ (Liu & Liu 2022, P.488). ➢ 3- تنقيب هيكل الويب.	

	➢ **يساعد على اكتشاف أوجه التشابه بين مواقع الويب أو اكتشاف مواقع مهمة لموضوع أو فرع معين أو في اكتشاف مجتمعات الويب. كما أنها تستخدم للكشف عن** **مخطط صفحة الويب، تعدين بنية الويب تبحث هذه العملية عن أوجه التشابه بين مواقع الويب أو كبيرة المواقع ذات الصلة ببحث المتعلم. Khudhair , & Ahmed ,2015, (P.104-105).** **وفي مصدر أخر، تعدين بنية الويب: تعدين بنية الويب هو تم تعريفها على أنها أداة تستخدم للعثور على جمعية تشارك في صفحات الويب ذات الصلة، إلى البيانات أو عن طريق الارتباط المباشر، •الخطوة 1: يتم تصنيف صفحات الويب تعتمد على الارتباط التشعبي والمعلومات، •الخطوة 2: العثور على بنية مستند الويب بحد ذاتها.** **• الخطوة 3: في موقع الويب الخاص بمجال معين طبيعة التسلسل الهرمي أو شبكة الارتباطات التشعبية (Shukla, et al.,2020,P.2)**

جدول (1): الفرق بين تنقيب البيانات وتنقيب الويب من إعداد الباحثة.

مناقشة دور التقنية في دعم إدارة المعرفة وعملياتها:

إن إدارة المعرفة تساعد الشركات على تحقيق تفوقها التنافسي، حيث توفر وسائل لإدارة واستخدام الموارد المعرفية بكفاءة، وإن توافق الآراء حول أهمية إدارة الموارد القيّمة داخل المؤسسات ضرورياً لضمان استدامتها وتحقيق المزايا التنافسية؛ لذا يتطلب التعامل مع المعرفة مجموعة من الأساليب بما في ذلك التنقيب عن البيانات لدعم عمليات إدارة المعرفة ، من هذا المنطلق تعتبر إدارة المعرفة (KM) مسألة أساسية في استراتيجية أعمال المؤسسات حيث تعد إدارة المعرفة عنصرًا حيويًا في المؤسسة سواءً تم اعتبارها من سمات البشر أو ملكية للمنظمة، لذلك تكون المعرفة بشكل صريح أو ضمني، وتتعدد طرق إدارة المعرفة داخل المؤسسات حيث كل منها يؤدي إلى تأثيرات مختلفة، وذلك من خلال إدارتها بشكل مناسب داخل البيئة التنظيمية، وإن تأثير الابتكار أو التكنولوجيا المعلوماتية على إدارة معرفة العاملين له دور مهم في تمكين المؤسسات من المشاركة في تقييم الفوائد الناتجة عن توظيف أنظمة مختلفة داعمة لإدارة المعرفة وعملياتها ، لذا فإن تعزيز ممارسات إدارة المعرفة بين الموظفين تعتبر حاجة ماسة إلى الفهم العميق للعوامل التي تؤثر بشكل كبير حيث تعتبر المعرفة أصول رئيسية تساهم في تحقيق نجاح الشركة مثل رأس المال والممتلكات والمواد والآلات (Othman& Others, 2019: P: 902).

التنقيب في البيانات في عملية إدارة المعرفة

إن عمليات استخراج البيانات لها دور مهم في اكتشاف معارف جديدة ضمن كميات هائلة من المعلومات، ويساعد تحليل البيانات الذكي ويشكل الأساس لإدارة المعرفة الفعالة. ومن خلال التمييز بين عمليتي اكتشاف المعرفة وتحليل البيانات الذكية، يمكن للمؤسسات تسخير كليهما لاستخراج المعرفة المفيدة من قواعد بيانات واسعة النطاق.

إن التحليل الذكي للبيانات في عمليات إدارة المعرفة يمتاز بقدرته على التنبؤ بالتطورات المستقبلية باستخدام أساليب المحاكاة وأنماط استخدام البيانات، كما يتيح تحليل مجموعات البيانات بناء النماذج التي توضح التفاعلات والخصائص الكامنة في مصادر البيانات المتنوعة حيث تخدم النماذج التنبؤية والوصفية أغراضًا مختلفة، تتنبأ النماذج التنبؤية بالنتائج بحسب مجموعات البيانات المعروفة بينما توضح النماذج الوصفية التبعيات والتفاعلات داخل مجموعات البيانات، ويهدف كلا النموذجين إلى الكشف عن الأنماط التي قد تتجاهلها الأساليب التقليدية؛ مما يساهم في فهم أعمق للأنظمة وتسهيل اتخاذ القرارات الصائبة.

يتم التعرف على بعض نموذج عمليات التنقيب عن البيانات في فرق المشروع من خلال الشكل (11) حيث تم تحديد نموذج لتقييم معرفة أعضاء فريق المشروع، باستخدام نظرية التثليث للجمع بين مصادر البيانات المتعددة وطرق الحصول على نتائج دقيقة، وتبدأ العملية بجمع البيانات من أعضاء الفريق بما في ذلك الدراسات الاستقصائية والمعلومات من المشاريع السابقة، ويقوم استخراج البيانات بتصنيف المعلومات

إلى قواعد بيانات خلفية معرفية وديموغرافية واستقصائية حيث تتضمن العمليات اللاحقة اكتشاف الأنماط والنمذجة التنبؤية والتفسير والتقييم لاستخلاص القواعد والأنماط لتنسيق تدفق البيانات والتحكم فيه داخل فرق المشروع.

يسهل هذا النموذج عملية اتخاذ القرار في المشروع والنجاح التنظيمي حول ديناميكيات الفريق وتوزيع المعرفة. ومن خلال فهم معارف أعضاء الفريق والاستفادة منها، يمكن للمؤسسات تحسين نتائج المشروع وتعزيز ثقافة التعلم والتحسين المستمر (Bushuyeva & Kutsenko,2019) .

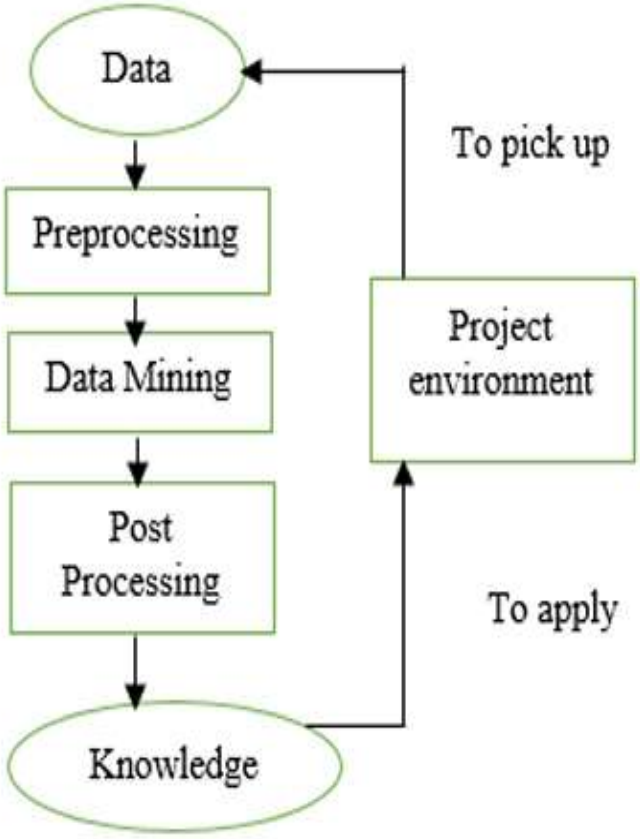


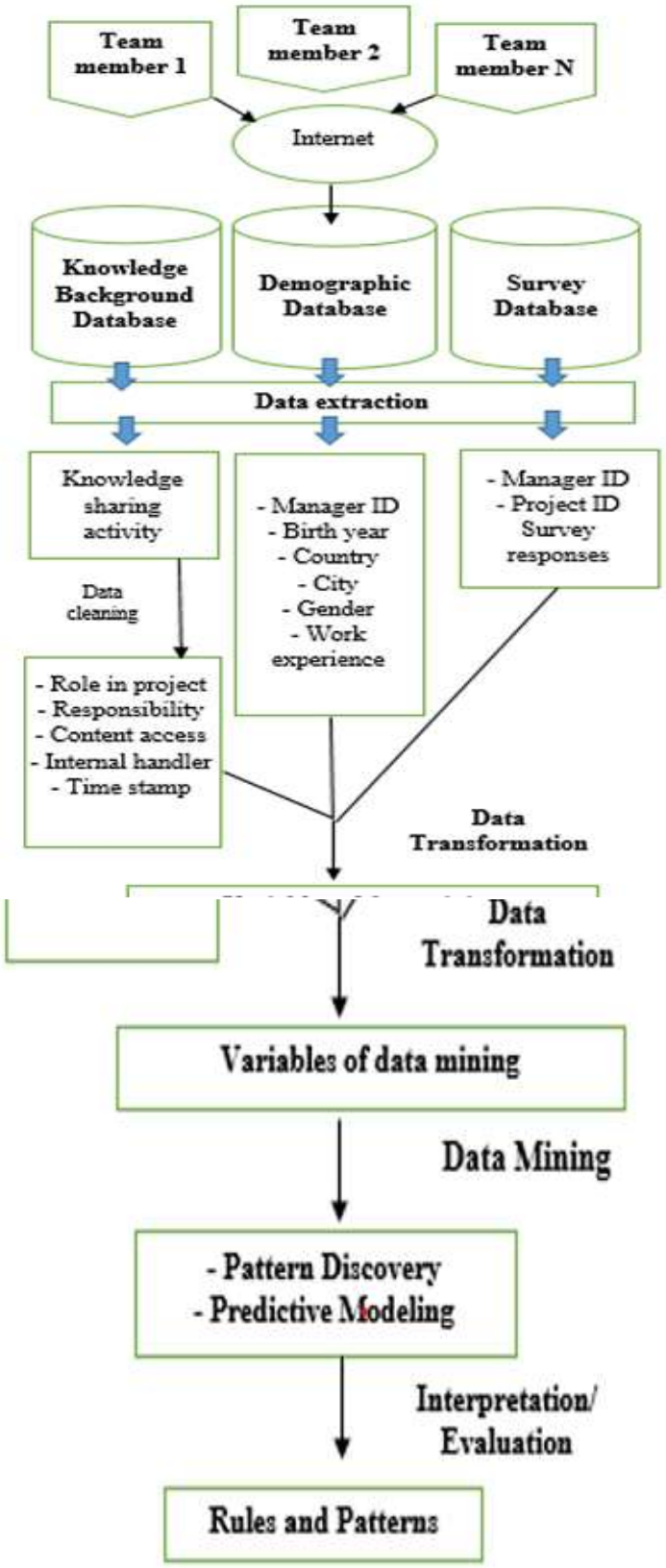
Fig. 3. Data Mining Cycle

الشكل (9) يوضح اكتشاف المعرفة في نموذج البيانات المختلط، (Bushuyeva & Kutsenko,2019,P:2-3)

Fig. 2 Knowledge Discovery in Data hybrid model

الشكل (10) يوضح دورة استخراج البيانات، (Bushuyeva & Kutsenko,2019,P:2-3)

Fig.1 Model of Data Mining Processes in project team



الشكل (11) يوضح نموذج عمليات التنقيب عن البيانات في فريق المشروع.

(Bushuyeva & Kutsenko,2019, P:2-3)

نتائج الدراسة:

1-توضح النتائج أنّ تطبيق خوارزميّات التنقيب على البيانات والويب يمكّن من اكتشاف علاقات وأنماط لم تكن ظاهرة بالطرق التقليديّة، الأمر الذي يدعم اتّخاذ قرارات أكثر دقّة في الوقت المناسب (Wang & Gao، 2019).

2-يسهم دمج مخرجات التنقيب في بوّابات أو مستودعات إدارة المعرفة في تقليل فقدان الخبرة، وتحسين مشاركة المعرفة بين العاملين، وزيادة الابتكار المؤسّسي (Othman وآخرون، 2019).

3-كشفت الدراسة عن تحدّيات مثل التعامل مع البيانات غير المهيكلة، ضعف التكامل بين المصادر، الاعتماد الكبير على الخبرة البشريّة، ومخاوف الخصوصيّة؛ إلّا أنّ اعتماد أطر منهجيّة مثل KDD وCRISP-DM، وتشكيل فرق عمل متعدّدة التخصّصات، يقلّل من آثار هذه المعوّقات (Dol & Jawandhiya، 2022؛ Martin، 2022).

4- برهنت دراسات حديثة على فعاليّة التكامل في التعليم عبر تخصيص المحتوى التعليمي (Liu & Liu، 2022)، وفي الرعاية الصحيّة بتحسين التشخيص المبكّر، وفي التسويق باستهداف العملاء بدقّة أكبر (Bhavana & Raheja، 2022).

أظهرت نتائج الدراسة أن المعرفة أداة ضرورية في المنظمات وفي بناء القدرة التنافسية واتخاذ القرار المستند إلى الأدلة. وقد أكدت النتائج أن **تقنية تنقيب البيانات** تمثل أحد المكونات الحيوية لأنظمة إدارة المعرفة، حيث ساهمت في تحويل البيانات الخام إلى معرفة قابلة للاستخدام العملي، تدعم مختلف وظائف المنظمة من تحليل وتخطيط وتنفيذ.

وقد كشفت الدراسة أن **عملية التنقيب عن البيانات تمر بعدة مراحل تكرارية** تشمل جمع البيانات، وتنظيفها، وتحليلها باستخدام خوارزميات معقدة، ثم تفسير النتائج. وتعتمد هذه العملية على أدوات تحليلية مثل التصنيف والتجميع وتحليل الارتباط، والتي بدورها تساعد في الكشف عن الأنماط الخفية في البيانات، خاصة في ظل تضاعف حجم البيانات الرقمية وتنوع مصادرها.

من أبرز المخرجات أن استخدام تقنيات التنقيب يسهم في تحقيق وفورات مالية واضحة من خلال خفض التكاليف التشغيلية، وتحسين استراتيجيات استهداف العملاء، وتقليل المخاطر، وتعظيم الإيرادات. كما تبين أن نجاح مشاريع التنقيب يرتبط بشكل مباشر بمهارات فريق العمل، حيث يتطلب توفر خبرات تقنية وتحليلية متنوعة، بالإضافة إلى القدرة على تحديد أهداف واضحة وتحويلها إلى مشكلات تحليلية قابلة للمعالجة.

وأوضحت الدراسة أن أدوات التنقيب تختلف من مشروع لآخر حسب طبيعة القطاع، وحجم البيانات، وتعقيد المشكلة وذلك يتطلب تقييم دقيق للمتطلبات قبل اختيار الأداة المناسبة. كما أكدت أهمية اتباع منهجيات معيارية مثل KDD و CRISP-DMلضمان اتساق خطوات التحليل وفعالية النتائج.

وعند ربط هذه الأدوات بإدارة المعرفة، يتضح أن **تكامل التنقيب مع نظم المعرفة** يعزز من جودة المحتوى المعرفي داخل المؤسسات، ويقلل من فقدان الخبرات، ويحفّز الابتكار، ويسرّع من تدفق المعرفة بين وحدات المنظمة. غير أن هذا التكامل لا يخلو من التحديات، خصوصًا فيما يتعلق بالتكاليف العالية، والحاجة إلى كوادر مؤهلة، إضافة إلى ضرورة معالجة قضايا الخصوصية وأمن البيانات.

توصيات الدراسة

1. التحول من الرؤية التقنية إلى الرؤية الاستراتيجية عند توظيف أدوات التنقيب، بحيث تخدم أهداف إدارة المعرفة لا مجرد تحليل البيانات.
2. الاستثمار في البنية التحتية التقنية والمؤسسية، بما في ذلك أدوات التحليل المتقدمة والأنظمة التي تضمن دمج نتائج التنقيب في مخازن المعرفة.
3. إعداد فرق تحليل متعددة التخصصات تضم خبراء في التنقيب، وإدارة المعرفة، والمجال التطبيقي المستهدف؛ بما يضمن التفسير الدقيق للنتائج.
4. إجراء دراسات حالة تطبيقية توضح الأثر العملي لتقنيات التنقيب على كفاءة إدارة المعرفة في قطاعات محددة كالتعليم أو الصحة أو الصناعة.
5. التأكيد على أهمية أمن المعلومات والامتثال للمعايير الأخلاقية عند التعامل مع البيانات الحساسة، وخاصة في بيئات التنقيب عن البيانات المرتبطة بالويب.

6. بناء ثقافة تحليلية داخل المؤسسة من خلال تدريب العاملين على فهم آليات التنقيب ومخرجاته واستخدام نتائجه في اتخاذ القرار.

مقترحات للبحوث المستقبلية

- تطوير نماذج هجينة تجمع بين تقنيات التعلم العميق والتنقيب المعرفي لتحسين جودة نتائج التحليل في الزمن الحقيقي.
- تحليل الأثر الاقتصادي المباشر لتكامل أدوات التنقيب مع إدارة المعرفة على مؤشرات الأداء المؤسسي.
- استكشاف دور التقنيات الناشئة مثل الذكاء الاصطناعي التوليدي، والحوسبة السحابية، والتحليلات التنبؤية في دعم أنظمة المعرفة الذكية.

الخاتمة

في ضوء التحوّل الرقمي المتسارع وما يصاحبه من تضخّم هائل في حجم البيانات وتنوّع مصادرها، تبيّن هذه الدراسة أنّ تكامل تقنيات تنقيب البيانات وتنقيب الويب مع نظم إدارة المعرفة يشكّل ركيزةً أساسيةً لتعظيم القيمة المعرفيّة داخل المؤسّسات. فالتنقيب يقدّم قدرة فائقة على استخراج الأنماط الخفيّة والتنبّؤ بالسلوكيّات، بينما توفّر إدارة المعرفة البنية التنظيميّة التي تضمن جمع تلك المعارف وتوثيقها وتوظيفها بفعاليّة. وعند دمج الطرفين، يصبح بالإمكان تحويل البيانات الخام إلى رأس مال معرفيّ يساهم في رفع كفاءة القرارات الاستراتيجيّة وتعزيز الميزة التنافسيّة المستدامة

(Shettigar وآخرون، 2022) (Fischer, & Dörpinghaus ,2024, P.53).

تعتبر هذه الدراسة من المواضيع الهامة في عصر البيانات والمعلومات، يتمثل في العلاقة التكاملية بين تقنيات تنقيب البيانات وتنقيب الويب من جهة، وإدارة المعرفة من جهة أخرى، وذلك في سياق يتسم بتزايد حجم البيانات، وتنوع مصادرها، وتعقيد احتياجات المؤسسات في استخلاص المعرفة الدقيقة لاتخاذ قرارات فعّالة ، وقد تم استعراض المفاهيم النظرية الأساسية لهذه المحاور، و تبين كيف تُمثل تقنيات التنقيب أدوات استراتيجية لاستخلاص الأنماط والمعلومات الخفية من البيانات الضخمة، وتحويلها إلى معرفة ذات قيمة داخلية قابلة للتوظيف ضمن أنظمة إدارة المعرفة. كما تم التطرق إلى أنواع تنقيب الويب (المحتوى، الاستخدام، الهيكل)، وأهم الخوارزميات والأدوات البرمجية المستخدمة، مثل WEKA و KNIMEوRapidMiner، ومدى إسهامها في دعم أداء المؤسسات في مجالات متعددة، كالتعليم، الصحة، والتسويق.

أظهرت الدراسة أن التكامل بين التنقيب وإدارة المعرفة يعزز من جودة القرارات المؤسسية، ويدعم الابتكار التنظيمي، إلا أنه يواجه تحديات فنية وبشرية وتنظيمية، من أبرزها صعوبة التعامل مع البيانات غير المهيكلة، وضعف التكامل بين المصادر، والاعتماد على خبرة بشرية متخصصة. وقد أوصت الدراسة بأهمية الاستثمار في البنية التحتية التقنية، وبناء فرق عمل متعددة التخصصات، وتبني نماذج منهجية مثل KDDو CRISP-DMلتوجيه عمليات التنقيب وتحقيق نتائج قابلة للتكرار.

في ضوء ما سبق، يمكن القول إن تنقيب البيانات وتنقيب الويب يشكلان ركيزة أساسية لتحويل البيانات إلى رأس مال معرفي فعّال، لكن الاستفادة القصوى منهما تتطلب التكامل بين التقنية والخبرة البشرية، وتطوير أطر تنظيمية مرنة تراعي التغيرات المستمرة في بيئات العمل الرقمية. وتفتح هذه الدراسة المجال أمام أبحاث مستقبلية يمكن أن تتناول بناء نماذج عملية لتكامل التنقيب مع إدارة المعرفة في بيئات واقعية، أو دراسة الأثر الاقتصادي لهذا التكامل على أداء المؤسسات.

إن النتائج هذه الدراسة توفر فرص جديدة أمام الباحثين والممارسين لاستثمار التحليل الذكي للبيانات في تعزيز رأس المال المعرفي وتحقيق التفوّق المؤسّسي المستدام.

المراجع:

المراجع العربية:

العازمي، ف. ب. ف. (2023). دور استخدام أسلوب التنقيب في البيانات في تحسين جودة عملية المراجعة ببيئة الأعمال الكويتية. المجلة العلمية للدراسات والبحوث المالية والإدارية، 15(عدد خاص)، 1-39.

- زعرور، إياد. (2015). التنقيب في البيانات والتوقعات الاقتصادية، إدارة المعهد الوطني، بيروت: لبنان.

المراجع الأجنبية:

Akulwar, P., Pardeshi, S., & Kamble, A. (2018, August). Survey on different data mining techniques for prediction. In *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*. https://doi.org/10.1109/i-smac.2018.8653734

Al-Rifai, S. S., Shaban, A. M., Shihab, M. S. M., Mustafa, A. S., Alhalboosi, H. A., & Shantaf, A. M. (2020, June). Paper review on data mining, components, and big data. In *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. https://doi.org/10.1109/hora49412.2020.9152919

Baban Jabbar Othman, Al-Kake, F., Mohd Diah, M. L., Othman, B., & Hasan, N. M. (2019). This study examines the antecedents and the effects of knowledge management and information technology in the manufacturing industry. *International Journal of Psychosocial Rehabilitation, 23*(2).

Beulah, R. (2019). A survey on different data mining techniques for crop yield prediction. *International Journal of Computer Sciences and Engineering, 7*(1), 738–744. https://doi.org/10.26438/ijcse/v7i1.738744

Bhavana, & Raheja, N. (2022, April 7). AERS: Adaptive and efficient hybrid recommendation system based on web usage mining. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*. https://doi.org/10.1109/icscds53736.2022.9760790

Bushuyeva, N., & Kutsenko, M. (2019, September). Data Mining technics in projects with multinational teams. *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT).* http://dx.doi.org/10.1109/stc-csit.2019.8929801

Dastyar, B., Kazemnejad, H., Sereshgi, A., & Jabalameli, M. (2017). Using Data Mining Techniques to Develop Knowledge Management in Organizations: A Review. *Journal of Engineering, Project, and Production Management*, 7, 80–89.

Dol, S. M., & Jawandhiya, P. M. (2022, July). Use of Data mining Tools in Educational Data Mining. *2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT).* http://dx.doi.org/10.1109/ccict56684.2022.00075

Er Kara, M., Oktay Fırat, S., & Ghadge, A. (2019). A data mining-based framework for supply chain risk management. *Computers & Industrial Engineering*, Accepted.

Fischer, A., & Dörpinghaus, J. (2024). Web mining of online resources for German labor market research and education: Finding the ground truth? *Knowledge*, 4(1), 51–67. https://doi.org/10.3390/knowledge4010003

Guo, H. (2021). Research on web data mining based on topic crawler. *Journal of Web Engineering*. https://doi.org/10.13052/jwe1540-9589.20411

Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann Publishers.

Jaiswal. (2024). KDD - Knowledge Discovery in Databases. Accessed April 20, 2024. https://www.javatpoint.com/kdd-process-in-data-mining

Jinyang, L. (2023, February 24). Application of data mining in logistics industry in the era of big data. *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA).* http://dx.doi.org/10.1109/eebda56825.2023.10090765

Khullar, H. (2023). What is Data Mining? Definition, Applications, Tools, Benefits, Process. Accessed March 31, 2024, from https://www.shine.com/blog/data-mining-guide

Khudhair, R., & Ahmed, A. (2015). COMBI: A NEW COMBINATION OF WEB MINING TECHNIQUES FOR E-LEARNING PERSONALIZATION. *International Journal of Technical Research and Applications*, 3, 2320–8163.

Liu, F., & Liu, J. (2022, July). A novel teaching model based on Web Mining and Data Mining. *2022 Global Conference on Robotics, Artificial Intelligence and Information Technology (GCRAIT).* http://dx.doi.org/10.1109/gcrait55928.2022.00107

Malviya, B. K., & Agrawal, J. (2015, April). A study on web usage mining theory and applications. *2015 Fifth International Conference on Communication Systems and Network Technologies.* http://dx.doi.org/10.1109/csnt.2015.247

Manamperi, A. (2023). The Role of Statistics in Data Mining.

Mardiani, Ermatita, & Samsuryadi. (2019). Data Mining Management for Implementation of Knowledge Management Using SECI Model and Data Testing. *International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, STMIK MDP Palembang, Indonesia.

Martin, Ostrovsky. (2022). 10 Key Data Mining Challenges in NLP and Their Solutions. Accessed April 18, 2024. https://www.dataversity.net/10-key-data-mining-challenges-in-nlp-and-their-solutions/

Md Selamat, S. A., Prakoonwit, S., & Khan, W. (2020). A review of data mining in knowledge management: applications/findings for transportation of small and medium enterprises. *SN Applied Sciences*, 2. https://doi.org/10.1007/s42452-020-2589-3

Sharma. (2024). Data Mining Architecture. Retrieved from https://www.upgrad.com/blog/data-mining-architecture

Shettigar, R., Bhise, Y., Bandgar, B., & Madhavedi, S. (2022). Data Mining Application in Effective Knowledge Management: An Empirical Study. *International Journal of Food and Nutritional Sciences*, 11, 2285–2296.

Shukla, R. K., Sharma, P., Samaiya, N., & Kherajani, M. (2020, February). Web Usage Mining - A Study of Web data pattern detecting methodologies and its applications in Data Mining. *2nd*

International Conference on Data, Engineering and Applications (IDEA). http://dx.doi.org/10.1109/idea49133.2020.9170690

Siregar, A., Siregar, Y., & Khairani, M. (2023). Implementation Of The Data Mining Cart Algorithm In The Characteristic Pattern Of New Student Admissions. *Journal of Computer Networks, Architecture and High Performance Computing*, 5, 263–275. https://doi.org/10.47709/cnahpc.v5i1.1975

Utkarsh. (2023). Copyright 2024 Interview Bit Technologies Pvt. Ltd.

Verma, P., & Kesswani, N. (2014). Web Usage mining framework for Data Cleaning and IP address Identification. https://www.researchgate.net/publication/26505245

Wang, A., & Gao, X. (2019). Multi-Tasks discovery method based on the concept network for data mining. *IEEE Access*, 7, 139537–139547. https://doi.org/10.1109/access.2019.2943462

Yu, K., Yuan, T., & Li, Y. (2021, January). Application of data mining technology in sports data analysis in colleges and universities. *2021 International Conference*