تمهيد:

إن الأنظمة التعليمية التي تعتمد بطبيعتها على الحاسوب، تملك القدرة على تسجيل كامل تفاصيل تفاعل الطالب مع النظام، ولما كانت تأثير هذه البيانات على دعم القرار المتخذ في المؤسسات التعليمية والأكاديمية كبير، ومع ظهور ثورة المعلومات والإنترنت التي اجتاحت معظم مناحي الحياة بشكل عام والمجال التعليمي بشكل خاص، وظهور العديد من الجامعات التي تعتمد بشكل كامل على الحاسوب والإنترنت، كانت الحاجة لتحليل البيانات الخام المجنية من هذه الأنظمة بهدف الوصول للمعرفة المخبئة في هذه البيانات وتحسين وضع الطلاب.

التنقيب عن المعرفة في المجال التعليميand Learning Analytics Educational Data Mining هو مجال بحثي جديد يُعرّف على أنه استخدام تقنيات التنقيب في البيانات الخام المستخرجة من الأنظمة والعمليات التعليمية بهدف مساعدة مصممي ومستخدمي هذه الأنظمة لتحسين أدائها واستخلاص معلومات مفيدة عن سلوك الطلاب في عملية التعلّم. كما أن استخدام تقنيات التنقيب في أنظمة التعلم الافتراضية يمكن أن يلعب دوراً هاماً وفعالاً في حل مشاكل التعلم وتقديم المعلومات التي تساعد أصحاب القرار من طلاب ومعلمين وواضعي الخطة الدراسية بأخذ قرارات مدعومة بالمعلومات عن كيفية التفاعل مع المصادر التعليمية.

وبذلك تكون عملية التنقيب السابقة تطبيق للأساليب الآلية بغية الكشف عن الأنماط في مجموعات كبيرة من البيانات التعليمية. قد يكون من الصعب أو المستحيل تحقيق ذلك نتيجة حجم البيانات الهائلة المستخلصة من الأنظمة التعليمية، إلا أن الجهود قد تكثفت في الآونة الأخيرة لإيجاد أدوات تسهل وتمهد الطريق أمام التنقيب بكمية هائلة من المعطيات وتعمل على تجهيز هذه المعطيات لمرحلة التنقيب قبل البدء بها.

# الفصل الأول: مقدمة عامة

## **مقدمة المشروع:**

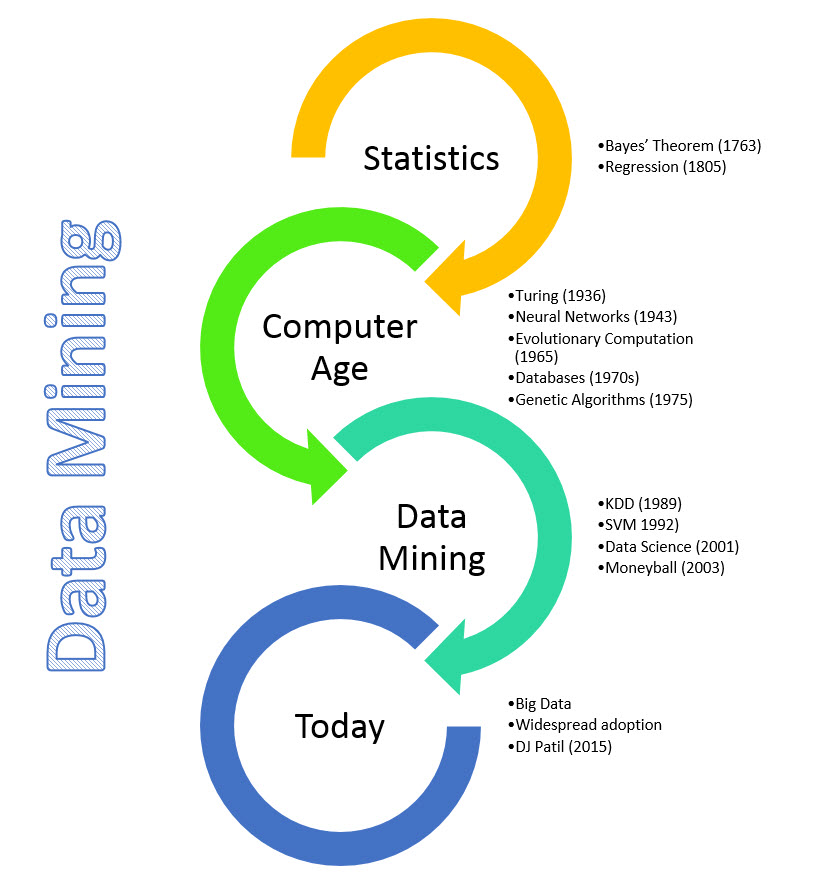
تتركز جهود الباحثين عند التنقيب عن المعرفة في الأنظمة التعليمة في محاولة إيجاد الأسئلة المتعلقة بكيفية التنبؤ بنجاح الطلاب، تسلسل السيناريو الأمثل والأكثر ملائمة لكل طالب في العملية التعليمة، ما هي المؤشرات المستوحاة من سلوك الطلاب والتي تدل على إمكانية نجاح أو فشل الطالب، ما هي صفات البيئة التعليمية الأمثل والتي تساهم إلى حد كبير في نجاح الطلاب.. وما إلى ذلك من الأسئلة التي تؤثر بشكل أو بآخر على الطلاب ونجاحهم أو فشل للاستفادة من أجوبة هذه الأسئلة وتكثيف الجهود لرفع مستوى الطالب ومساعدته وتخفيض نسبة الفشل لدى هؤلاء الطلاب.

إنّ مصدر الداتا التي يتم تطبيق عمليات التنقيب عليها لا يقتصر على البيانات الناتجة عن الحالات الفردية للطلاب (كسلوك الملاحة في النظام، التمارين التفاعلية والامتحانات)، وإنما يشمل أيضاً البيانات الناتجة من تعاون الطلاب، البيانات الإدارية، أو حتى البيانات الديموغرافية (كالعمر، الجنس..). يمكن أيضاً الاستفادة من البيانات التي تعبر عن تأثر الطلاب والحالات العاطفية لهم والتي يمكن استنتاجها من التعابير المستخدمة وغيرها.

أما عن كمية الداتا التي عادة ما يتم استخدامها في عملية التنقيب فهي ضخمة جداً، فمع تزايد استخدام أجهزة المحمول والتقنيات الـ online في العمليات التعليمية، أتيحت كمية هائلة من الداتا تِبعاً لإمكانية الحصول على نتيجة كامل العمليات وأشكال تفاعل الطالب مع المحتوى التعليمي وتسجيلها “logging” وجعلها متاحة للأغراض التحليلية. كما أنه أصبح من المتوفر الآن كميات كبيرة من البيانات للباحثين من خلال السجلات العامة مثل مركز “Pittsburgh Science of Learning Center Datashop”.

## لمحة تاريخية:

علم التنقيب في المعطيات Data Mining ليس بجديد، وخاصة نتيجة اعتماده بشكل أساسي على علم الإحصاء الذي تعود أصوله إلى منتصف القرن الثامن عشر. فالوضع الحالي الذي آل إليه التنقيب عن المعطيات جاء نتيجة جهود تراكمت من سنين عديدة، إلا أن هذا المجال شهد تطوراً ملحوظ في الآونة الأخيرة. يبين الشكل التالي مراحل تطور هذا المجال



الشكل1Data Mining Timeline

بالعودة إلى جذور التنقيب عن المعرفة في الأنظمة التعلمية نجد أن الأبحاث في هذا المجال جديدة نسبياً. فبالرغم من أن عمليات البحث والتحليل للبيانات الناتجة عن البرمجيات التعلمية بدأت من مدة طويلة، إلا أن التنقيب عن المعرفة في الأنظمة التعليمية Educational Data Mining (EDM) تم اعتباره كحقل مستقل بحد ذاته مؤخراً، من خلال المؤتمرات (Internal Conference on Educational Data Mining, started in 2008) والمجلات العلمية (Journal of Educational Data Mining “JEDM”, first issue published 2009). كما تم تقديمها وتعزيزها في مؤتمرات تكنولوجيا التعليم مثل:

the International Conference on Educational Data Mining, the Conference on Learning Analytics and Knowledge, the International Conference on Artificial Intelligence in Education (AIED), the International Conference on Intelligent Tutoring Systems (ITS) and the International Conference on User Modeling, Adaptation, and Personalization (UMAP).

## التنقيب في البيانات:

تّعنى التنقيب في البيانات بتحليل هذه البيانات من أُطر ونواحي وأبعاد مختلفة بهدف الوصول للأنماط المخبأة في مجموعة كبيرة ومعقدة من المعطيات Big Data، مما جعل البعض يطلق عليها اكتشاف المعرفة Knowledge discovery. تلعب عملية التنقيب هذه دوراً هام في عمليات دعم القرار وتسمح بالقيام باستعلامات دقيقة باستمرار دون الحاجة إلى وجود هدف محدد. ولا تقوم عملية التنقيب عن المعطيات على التصفح العشوائي لقواعد المعطيات وإنما تكون عملية مدروسة بدقة تامة ومصممة لاستخلاص نتائج محددة وقرارات مفيدة غنية بالمعلومات المهمة والمطلوبة. تقوم عملية التنقيب هذه على بيانات ضخمة متراكمة على مدى سنين لتحقيق نتائج أدق وأفضل. تعمل أنظمة التنقيب على الحصول على معرفة ناتجة عن التنقيب في بيانات ناتجة عن أنظمة مختلفة بعد إجراء عمليات مختلفة عليها بغية مكاملتها والتأكد من صحتها، وهنا يجب الانتباه لأهمية إجراء عمليات التنقيب على بيانات صحيحة وخالية من الأخطاء قدر المستطاع، لأن إجراء التنقيب على بيانات خاطئة يؤدي للوصول لمعرفة خاطئة وغير دقيقة.

## مجالات التنقيب في البيانات:

### التنقيب في البيانات في المجال الطبي:

تحوي قواعد معطيات المشافي والمراكز الصحية كمية كبيرة من البيانات حول المرضى ووضعهم الصحي، وبدراسة العلاقات والأنماط بين هذه البيانات يمكن الوصول إلى معرفة طبية جديدة تفيد في اكتشاف مبكر للأمراض، مساعدة بتشخيص الأمراض. كما أن عملية التنقيب هذه تعمل على ضمان تلقي المريض العلاج المناسب في الزمان والمكان المناسبين. من أحد أهم تطبيقات التنقيب في المجال الطبي الضمان الصحي واكتشاف حالات الاحتيال. يعتقد الباحثون أن التنقيب في المعطيات الطبية يمكن أن يحسّن المستوى الصحي العام بنسبة تصل إلى 30%.

### التنقيب في البيانات في مجال القطاع المصرفي والمالي:

تستفيد المؤسسات المالية الضخمة كالبنوك من عمليات التنقيب في اكتشاف الربح والخسارة، واكتشاف الاحتيال، ودراسة الحملات التسويقية لاتخاذ القرار الأمثل اتجاه زبائنها للمحافظة عليهم وزيادة عددهم تبعاً لنوع الخدمات المصرفية المتنوعة التي تقدمها من قروض واستثمار وغير ذلك. تبرز أهمية التنقيب في المعطيات المالية نظراً لكمية البيانات الهائلة التي تُولّد بشكل سريع جداً مما يمنع الخبراء من المتابعة الدائمة والدقيقة لهذ المعطيات. أما عن مجال المبيعات فيعتمد التنقيب فيه على ربط شراء منتج من المنتجات بشراء منتج آخر. تسمح هذه العملية للتاجر بفهم سلوك الزبائن ومعرفة احتياجاتهم للعمل على تلبيتها وزيادة الأرباح. كما تساعد التجار في دراسة المنافسين لوضع خطط المبيعات الأمثل للتغلب على المنافسين.

### التنقيب في البيانات في مجالات العلوم والهندسة:

تستخدم التنقيب في البيانات لتحليل الصور الفلكية والهندسة الكيميائية والمجال الحيوي. كما يمكن الاستفادة من التنقيب في مجال الأبحاث. إذ تقدم عمليات التنقيب للباحثين القدرة على إيجاد بيانات مماثلة والتي قد تسهل أو تغيير مجرى البحث.

### التنقيب في البيانات في المؤسسات التعليمية:

يوجد عدة تعريفات للتنقيب في البيانات التعليمية من بينها تعريف ذُكر في مجلةEducational Data Mining و Baker الذي يعد من أوائل الباحثين في التنقيب بالمجال التعليمي بأنها تطوير الأساليب لاستكشاف الأنواع الفريدة من البيانات التي تأتي من البيئات التعليمية، واستخدام تلك الأساليب من أجل فهم أفضل للطلاب، وعكسها على البيئة التي يتعلمون بها.

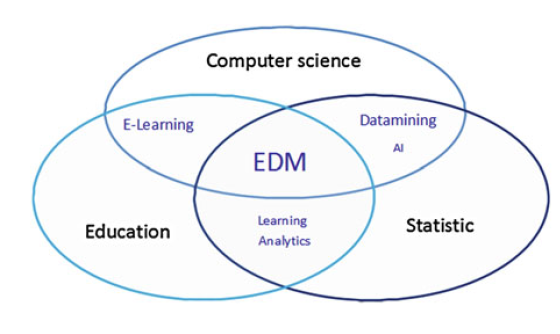
وفي تعريف آخر: استخدام تقنيات التنقيب عن المعلومات لأنواع محددة من البيانات المستخرجة من البيئات التعليمية لمعالج وحل المسائل والمشاكل التعليمية الهامة.

كلا التعريفين يؤكدان على استخلاص المعرفة من بيانات تعليمية لتحسين الأنظمة التعليمة.

عادة ما يتم الخلط بين عملية التنقيب السابقة وتحليل التعلم ‘learning analytics’ والذي يعرّف بجمع وتحليل البيانات وإعداد التقارير حول المتعلمين بغرض فهم وتحسين العملية التعليمية والبيئة التي تحضنها. ومع عدم وجود جدار فصل حقيقي بين المجالين ووجود العديد من الأهداف المشتركة بينهما تزداد صعوبة الفصل وتميز كون تطبيق يتبع لأحد المجالين أو للآخر.

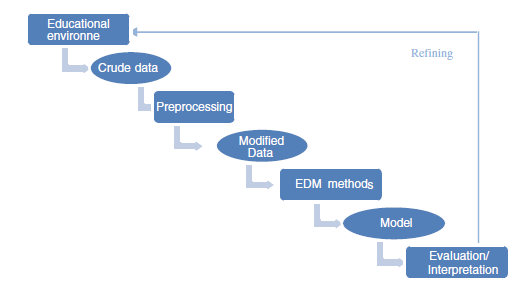
يمكن اعتبار التنقيب عن المعرفة في المجال التعليمي مؤلّف من ثلاث مكونات رئيسية: علوم الحاسب computer science، التعليم education، والإحصائيات statistics. تقاطع المجالات الثلاثة السابقة يشكل مجالات أخرى فرعية مثل تحليل التعليم Learning Analytics، التعليم الإلكتروني E-Learning، والنقيب Datamining

يُعد التنقيب في البيانات التعليمية مجال متعدد الاختصاصات، إذ يستخدم طرق وتقنيات مختلفة مثل الإحصاء، التنقيب في المعطيات، نظم استرجاع المعلومات، نظم الاقتراح، تعلّم الآلة، علوم التربية وأصول التدريس وغيرها. إلا أن عملية اختيار التقنيات المستخدمة تعتمد بشكل أساسي على القضية التعليمية المطروحة.



الشكل2 مجالات يرتبط التنقيب في البيانات التعليمية بها

يمكن تمثيل المراحل التي تمر فيها عملية التنقيب في الأنظمة التعليمية بالشكل التالي:



الشكل3 مراحل التنقيب في المعطيات التعليمية

هناك مجالات أخرى يتم تطبيق التنقيب عن المعطيات فيها مثل التحقيق بالجرائم والهندسة الصناعية واكتشاف الاحتيال وإدارة العلاقة بالعملاء وغيرها.

## أهداف التنقيب عن المعطيات في الأنظمة التعليمية:

تم تصنيف أهداف التنقيب في الأنظمة التعليمة بناء على أنواع المستخدمين النهائيين المستفيدين من المعرفة الناتجة عن التنقيب إلى:

* متعلمين: تقديم الملاحظات والنصائح للطلاب، الاستجابة لاحتياجات الطالب، العمل على رفع إنتاجية الطالب وسويته العملية.. إلخ
* معلمين: مساعدتهم على الفهم الأمثل لطبيعة تلاميذهم وكيفية سير العملية التعليمية ومستوى الطلاب وعكس ذلك على الطرق التي يتبعونها في التعليم وتحسين الأداء التدريسي.. إلخ
* باحثين: تطوير ومقارنة تقنيات استخراج البيانات لجعلها قادرة على تقديم الاقتراحات الأمثل لكل مشكلة أو مهمة تربوية محددة. تقييم فاعلية التعلم عند استخدام طرق وبيئات وأساليب مختلفة.. الخ
* إداريين: إيجاد الطريقة الأفضل لتنظيم المصادر (البشرية والمادية) وتقييم مؤسساتهم التعليمية.

في وجه النظر السابقة تم تصنيف الأهداف تبعاً لأنواع المستخدمين النهائيين، ولكنه من الصعب تصنيف جميع أهداف التنقيب عن المعرفة في الأنظمة التعلمية تبعاً لأنواع المستخدمين، وخاصة عند وجود أهداف ترتبط بأكثر من مستخدم، لذلك وجدت أبحاث أخرى قامت بتصنيف أهداف تنقيب البيانات التعليمية كالتالي:

* نمذجة الطالب: نمذجة المستخدم في النظام التعليمي تتطلب إعطاء معلومات مفصلة عن خصائص الطلاب وحالاتهم كالمعرفة التي يمتلكها الطالب والمهارات والخبرات والدوافع للتعلم وصولاً إلى دراسة بعض أنواع المشاكل التي قد تؤثر سلباً على مخرجات العملية التعليمية كاستخدام مصادر تعليمية غير فعالة. الهدف هنا هو خلق أو تحسين نموذج الطالب من انطلاقاً المعلومات المتاحة.
* توقع مستوى الطلاب التعليمي ونتائجهم: الهدف هو توقع النتيجة النهائية، وتوقع نتيجة أي عملية تعليمية انطلاقاً من بيانات ناتجة عن تفاعل الطالب مع المادة الدراسية.
* إيجاد الاقتراحات للطلاب: فاهدف هو اقتراح المحتوى أو حتى المهام الأمثل لوضع الطالب في الوقت الحالي.
* تحليل سلوك المتعلم: وهذا الهدف يأخذ أشكال عدّد، بدء من الإجابة عن الأسئلة الثلاثة سابقة الذكر، وصولاً لتصنيف الطلاب تبعاً لعوامل عدة كملفهم الشخصي.
* التواصل مع الـ stockholders: الهدف هو مساعدة المعلمين ومديري المواد للتحليل أنشطة الطلاب.
* تحليل هيكلية المجال: والهدف هو تحديد بنية المجال لمعرفة المحتوى الأمثل الذي يمكن إضافته للمواد والتسلسل الأفضل للعملية التعليمية، وذلك باستخدام القدرة على التنبؤ بأداء الطلاب وتحديد المحتوى والبيئة التعليمية المثلى لهم.
* تحسين المحتوى التعليمي: وهذا ذو ارتباط وثيق بالهدفين السابقين. فالهدف هنا هو تحديد كيفية تحسين المواد التعليمية (محتويات، أنشطة وما إلى ذلك) باستخدام معلومات حول الطالب.
* دراسة تأثيرات الدعم التربوي pedagogical support الذي يتم تقديمه من قبل أنظمة التعليم.

من الجدير بالذكر أن أهداف تنقيب البيانات التعليمية سابقة الذكر تهدف إلى تحسين النظم التعليمية بشكل عام والنظم التعليمية المعتمدة على الحاسوب بشكل خاص، يمكن القول بأن نمذجة الطالب هو المفتاح الأساسي لتحقيق العديد من أهداف ومهام تنقيب البيانات التعليمية.

## الطرق المستخدمة في تنقيب البيانات التعليمية:

لتحقيق أهداف تنقيب البيانات التعليمية تم استخدام معظم طرائق تنقيب البيانات التقليدية (clustering, classification, association analysis..) والتي من الممكن أن تصنف كالتالي:

* **التوقع Prediction:** يكون هدف هذه الطريقة هو تطوير نموذج يؤدي إلى استدلال aspect معين من هذه البيانات (يطلق عليه Predicate Variable) من مجموعة من الـ aspects من هذه البيانات (يطلق عليها Predictor Variables). من بعض طرق التوقع هذهClassification, Regression Density Estimation,. مثال على تطبيقات تنقيب البيانات التعليمية باستخدام هذه الطريقة هي توقع نجاح الطالب أكاديمياً وتوقع سلوكه.
* **العنقدة Clustering**: تعني باختصار تصنيف العينات التي تحويها البيانات والتي تتشابه فيما بينها بشكل طبيعي (عمليّاً يتم استخدام بعض طرائق قياس البعد بين بعضها البعض لتحديد مقدار التشابه) وعند تحديد هذه التصنيفات من الممكن معرفة انتماء عينة جديدة ما وفقاً لهذه التصنيفات الأخيرة. تستخدم هذه الطريقة في تنقيب البيانات لتجميع الطلاب وفقاً لنمط تعلمهم أو استراتيجية cognitive.
* **تنقيب العلاقات Relationship Mining:** تستخدم لاكتشاف علاقات بين المتغيرات في عينة من البيانات، يوجد عدّة أنواع من تنقيب العلاقات مثل association rule mining, sequential pattern mining, correlation mining casual data mining,. في تنقيب البيانات التعليمية يستخدم تنقيب العلاقات لتحديد العلاقات بين نشاطات الطلاب على الشبكة وعلاماتهم مثلاً.
* **البيانات للتحكيم البشري Distillation:** تتضمن هذه الطريقة depicting للبيانات بشكل يسمح للمستخدم بالتعرف أو تصنيف ميزات البيانات بشكل سريع وسهل الاستخدام. تعتمد هذه التقنية على تلخيص البيانات Summarization وتصويرها Visualization واستخدام الواجهات التفاعلية Interactive Interfaces لتوضيح المعلومات المفيدة ودعم اتخاذ القرار.
* **الاستكشاف مع النماذج Discovery With Models:** تقوم هذه الطريقة باستخدام نموذج معرّف ومحقق مسبقاً.
* **اكتشاف العينات الشاذة Outliers Detection**: تعتمد هذه الطريقة على اكتشاف عينات معينة من البيانات ذات اختلاف ملحوظ عن باقي العينات. تستخدم هذه الطريقة في تنقيب البيانات التعليمية لكشف الاختلافات بأفعال المتعلم أو المدرس وسلوكه، عمليات التعليم غير النظامية أو تحديد المتعلمين الذين يعانون من صعوبات في عملية التعلم.
* **تحليل الشبكات الاجتماعية Social Network Analysis:** يهدف التحليل الهيكلي SNA إلى دراسة العلاقة بين الأفراد بدل من دراس سلوك وخصائص الأفراد أنفسهم، إذ يتم تمثيل العلاقات الاجتماعية بعقد تمثل الأفراد وروابط بين هذه العقد تمثل العلاقات بين هؤلاء الأفراد من صداقات وعلاقات وتعاونية وغيرها. يمكن استخدام ذلك في تنقيب البيانات التعليمية لتفسير وتحليل العلاقات في المهام التعاونية والتفاعل مع أدوات التواصل.
* **تنقيب العملية Process Mining**: تهدف إلى استخلاص العمليات المتعلقة بالمعرفة، وذلك باستخدام سجلات الأحداث التي سُجلت من قبل نظام المعلومات لتكوين تمثيل مرئي واضح لكامل العملية. تتكون من ثلاثة حقول فرعية: التوافق conformance، التحقق checking، واكتشاف النموذج model discovery. في نقيب البيانات التعلمية يمكن استخدام تنقيب العملية لعكس سلوك الطلاب من حيث تأثير دراستهم (المكونة من تسلسل صفوف، علامات... إلخ) لكل طالب.
* **تنقيب في النصوص Text Mining:** يركّز على إيجاد واستخراج أنماط مفيدة، نماذج، وقواعد من النصوص غير المهيكلة مثل ملفات HTML، رسائل الدردشة ورسائل البريد الإلكتروني. تشمل المهام تصنيف النص، تجميع النص، تحليل للمشاعر المضمنة في النص، تلخيص للوثائق، نمذجة علاقات الكائنات وغيرها. إذ يستخدم التنقيب في النصوص لتحليل مضمون المناقشات، المنتديات، الدردشات، صفحات الويب، الوثائق.. الخ

## البيانات المحللة:

يوجد العديد من أنواع البيانات التي يتم استخدامها في عملية تنقيب البيانات التعليمية وذلك حسب الهدف من عملية التنقيب والتقنيات المستخدمة ويمكن تصنيف هذه البيانات وفق الخواص التالية:

* **وفرة البيانات Data Availability:**

مثل البيانات المسجلة على مدار السنين في قواعد معطيات المؤسسات التعليمية أو من سجلات النظم البرمجية المستخدمة.

* **مصادر التجميع Collection Resources:**

أي كان تجميع هذا البيانات قد تم بشكل يدوي من قبل موظفين ومراقبين في المؤسسة التعليمية أو بيانات رقمية قد تجمعها على شكل أرقام (علامات) أو سجلات في قواعد معطيات أو ملفات صوتية أو أشرطة فيديو

* **بيئة التعلم Learning Environment:**

إذ يتم تجميع البيانات الناتجة عن عملية التعلم في البيئة الافتراضية Logging من ملاحة في النظام وعدد نقرات و.. والاستفادة من ذلك في الوصول إلى معرفة تساعد في اتخاذ قرار يصب في مصلحة الطالب.

## بعض الأدوات المستخدمة في عملية التحليل:

من أهم التطوّرات التي شهدها هذا المجال هو إيجاد طرق لإدارة الكميات الضخمة من البيانات مما يسهل إجراء عمليات التحليل عليها. فقط طورت غوغل MapReduce للحد من التحديات التي تظهر نتيجة إدارة البيانات على نطاق الإنترنت، بما في ذلك توزيع البيانات والتطبيقات التي تمتلك بيانات مرتبطة بشبكات الحواسيب. فقواعد البيانات السابقة كانت غير قادرة على إدارة البيانات على نطاق شبكة الإنترنت، مما أدى إلى تطوير Apache Hadoop الذي يستخدم الآن بشكل شائع في إدارة البيانات.

بالإضافة للأدوات المستخدمة في إدارة البيانات ظهرت العديد من الأدوات التي تدعم عمليات التحليل وتؤمن عمليات التنقيب عن المعرفة دون الحاجة لمعرفة عميقة في علوم الحاسب، معظم هذه الأدوات مطور بما يتوافق مع مجال ذكاء الأعمال والذي يتجلى في ظهور SAS وأدوات IBM في المجال التعليمي، والتي استخدمت في التحليلات التنبؤية وتحسين عملية صنع القرار من خلال تحليل كميات كبيرة من البيانات وعرضها بشكل تفاعلي. فيما سبق كان استخدام الأدوات التحليلية عملية معقدة تتطلب فهم عميق في التحليل والبرمجة، أما الآن فحتى الأدوات المعقدة كـ SAS, RapidMiner and SPSS تتطلب معرفة تقنية أقل. وهنا لا بد من الإشارة أن كثير من برامج الحاسوب مثل Microsoft Excel شهد تطور ملحوظ في التحليل والإظهار. أدوات أخرى مثل Tableau Software تم تصميميها لدعم استخدام أدوات التحليل دون الحاجة إلى معرفة تقنية عالية، مما أدى إلى انتشار عمليات الـ EDM, LA بين عدد كبير من الباحثين في مجال علوم التعليم.

# الفصل الثاني: الدراسة المرجعية.