# تمهيد:

إن الأنظمة التعليمية التي تعتمد بطبيعتها على الحاسوب، تمتلك 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.

الفهرس:

[تمهيد: 1](#_Toc476440146)

[الفصل الأول: مقدمة عامة 3](#_Toc476440147)

[مقدمة المشروع: 4](#_Toc476440148)

[لمحة تاريخية: 5](#_Toc476440149)

[التنقيب في البيانات: 6](#_Toc476440150)

[مجالات التنقيب في البيانات: 7](#_Toc476440151)

[التنقيب في البيانات في المجال الطبي: 7](#_Toc476440152)

[التنقيب في البيانات في مجال القطاع المصرفي والمالي: 7](#_Toc476440153)

[التنقيب في البيانات في مجالات العلوم والهندسة: 7](#_Toc476440154)

[التنقيب في البيانات في المؤسسات التعليمية: 8](#_Toc476440155)

[أهداف التنقيب عن المعطيات في الأنظمة التعليمية: 10](#_Toc476440156)

[الطرق المستخدمة في تنقيب البيانات التعليمية: 12](#_Toc476440157)

[تصنيف البيانات المحللة: 14](#_Toc476440158)

[بعض الأدوات المستخدمة في عملية التحليل: 15](#_Toc476440159)

[الفصل الثاني: الدراسة المرجعية. 17](#_Toc476440160)

[مقدمة: 18](#_Toc476440161)

[بعض الأنظمة المشابهة: 19](#_Toc476440162)

[Open University Analyse: 19](#_Toc476440163)

[Completing the Loop: 23](#_Toc476440164)

[الفصل الثالث: الدراسة التحليلية. 26](#_Toc476440165)

[أنواع البيئات الافتراضية التعليمية: 27](#_Toc476440166)

[Massive Open Online Courses (MOOC): 27](#_Toc476440167)

[Intelligent Tutoring Systems: 27](#_Toc476440168)

[Adaptive and Intelligent Hypermedia Systems: 28](#_Toc476440169)

[Test and Quiz Systems: 28](#_Toc476440170)

[Learning Management Systems: 28](#_Toc476440171)

[أنواع أخرى للأنظمة التعليمية: 31](#_Toc476440172)

[أنواع البيانات: 32](#_Toc476440173)

[عملية ما قبل المعالجة: 32](#_Toc476440174)

# الفصل الأول: مقدمة عامة

## **مقدمة المشروع:**

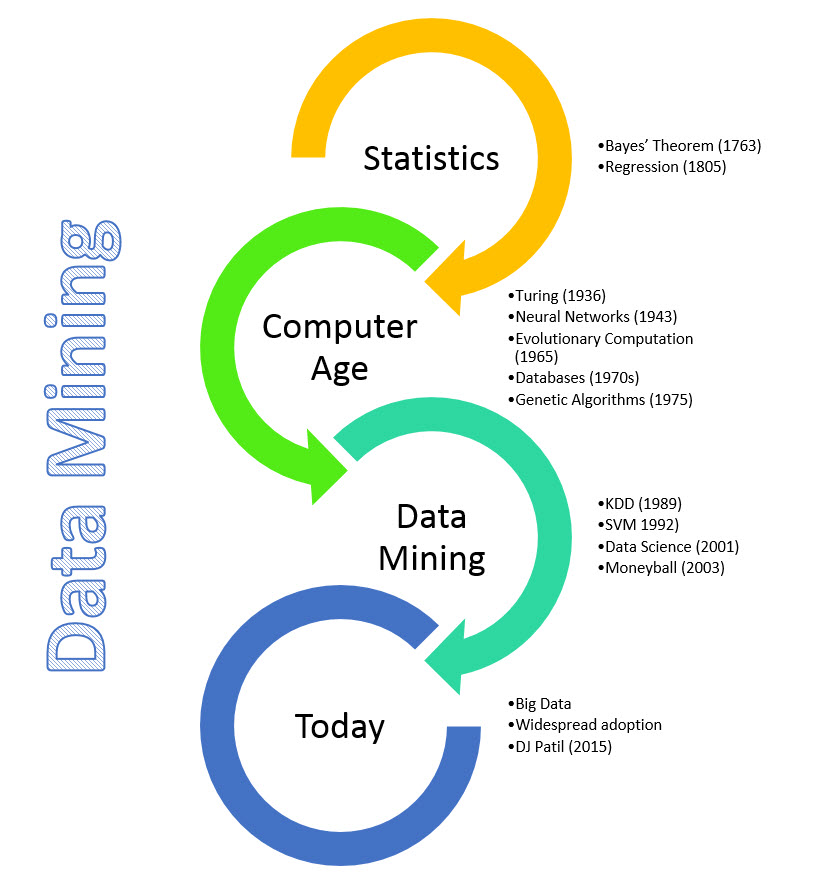
تتركز جهود الباحثين عند التنقيب عن المعرفة في الأنظمة التعليمية في محاولة إيجاد الأسئلة المتعلقة بكيفية التنبؤ بنجاح الطلاب، تسلسل السيناريو الأمثل والأكثر ملائمة لكل طالب في العملية التعليمية، ما هي المؤشرات المستوحاة من سلوك الطلاب والتي تدل على إمكانية نجاح أو فشل الطالب، ما هي صفات البيئة التعليمية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

إنّ مصدر البيانات 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 البيانات التي عادة ما يتم استخدامها في عملية التنقيب فهي ضخمة جداً، فمع تزايد استخدام أجهزة المحمول والتقنيات الـ online في العمليات التعليمية، أُتيحت كمية هائلة من البيانات 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)
* 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 باستخدام خوارزميات رياضية وحسابية، 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 في المعطيات على التصفح العشوائي لقواعد المعطيات وإنما تكون عملية مدروسة بدقة تامة ومصممة لاستخلاص نتائج وقرارات مفيدة غنية بالمعلومات المهمة والمطلوبة، بالرغم من ذلك تتم عملية التنقيب دون فرضيات مسبقة. تقوم عملية التنقيب هذه على بيانات ضخمة متراكمة على مدى سنين لتحقيق نتائج أدق وأفضل. وتعمل 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، والمساعدة 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 للتجار بفهم سلوك الزبائن ومعرفة احتياجاتهم للعمل على تلبيتها وزيادة الأرباح. كما تساعدهم 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.

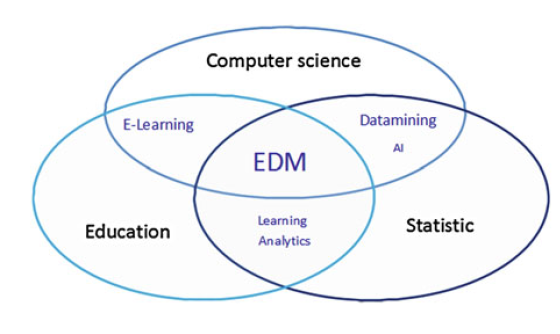
وفي تعريف آخر: استخدام تقنيات التنقيب عن المعلومات لأنواع محددة من البيانات المستخرجة من البيئات التعليمية لمعالج وحل المسائل والمشاكل التعليمية الهامة.

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

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. تُعنى العملية السابقة بالإجابة على أسئلة *محددة*. مع عدم وجود جدار فصل حقيقي بين المجالين ووجود العديد من الأهداف المشتركة بينهما تزداد صعوبة الفصل وتميز كون تطبيق يتبع لأحد المجالين أو للآخر.

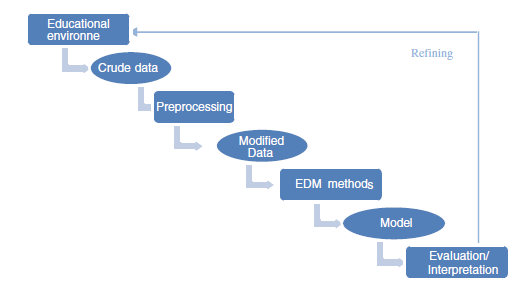
يمكن اعتبار التنقيب عن المعرفة في المجال التعليمي مؤلّف من ثلاث مكونات رئيسية: علوم الحاسب computer science، التعليم education، والإحصائيات statistics. تقاطع المجالات الثلاثة السابقة يشكل مجالات أخرى فرعية مثل تحليل التعليم Learning Analytics، التعليم الإلكتروني E-Learning، والنقيب Datamining

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

على الرغم من أن الأساليب المستخدمة للتنقيب في المحتوى الناتج عن الأنظمة التعليمية يتشابه مع مجالات أخرى كالتجارة الإلكترونية إلا أن الأهداف تكون مختلفة. فمن وجهة نظر النظام، لا توجد فروق لأن الهدف من التنقيب على شبكة الإنترنت في كلا المجالين هو دراسة سلوك المستخدمين (العملاء في مجال التجارة الإلكترونية، والطلاب في أنظمة التعلم الإلكتروني)، تقييم هذا السلوك وتحسين الأنظمة لمساعدة المستخدمين. إلا أنه من وجهة المستخدم هناك اختلافات، إذ أن التجارة الإلكترونية تهدف إلى توجيه العملاء في عملية الشراء في حين أن هدف التعليم الإلكتروني هو توجيه الطلاب في التعلم. يمكن تلخيص ما سبق بأن الفروق الرئيسية بين التنقيب في البيانات التعليمية والمجالات الأخرى هو: نوعية البيانات، الهدف والتقنيات نظراً لخصوصية بنية الأنظمة 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 التعليمية 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 انطلاقاً من المعلومات المتاحة.
* توقع المستوى التعليمي للطلاب ونتائجهم: الهدف هو توقع النتيجة النهائية، وتوقع نتيجة أي عملية تعليمية انطلاقاً من بيانات ناتجة عن تفاعل الطالب مع المادة الدراسية.
* إيجاد الاقتراحات للطلاب: فالهدف هو اقتراح المحتوى أو حتى المهام الأمثل لوضع الطالب في الوقت الحالي.
* تحليل سلوك المتعلم: وهذا الهدف يأخذ أشكال عدّة، 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

## تصنيف 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 كون تجميع هذا البيانات قد تم بشكل يدوي من قبل موظفين ومراقبين في المؤسسة التعليمية أو أنها بيانات رقمية قد تم تجميعها 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 الملاحة في النظام وعدد نقرات.. إلخ، 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.

ظهر مؤخراً أدى جديدة تدعى Apache Spark الذي يتمتع بميزات عديدة تجعله يتفوق على الأدوات الأخرى كمعالجة البيانات المتدفقة real-time processing

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 بظهور 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, DBMiner, DB2 Intellegent Miner, SPSS تتطلب معرفة تقنية أقل إلا أن معظمها تجاري(غير مجاني). 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.

يوجد برمجيات أخرى مفتوحة المصدر تؤمن عمليات التنقيب مثل Weka, Keel. كما يوجد بعض البرمجيات المختصة بالتنقيب عن المعطيات التعليمية مثل: MultiStar الذي يعمل على التصنيف ودراسة علاقات الارتباط، EPRules لدراسة علاقات الارتباط، KAON للعنقدة والتنقيب في النصوص، Synergo/ColAT للحصول على الإحصائيات وعمليات عرض النتائج، GISMO للعرض، Listen tool للعرض، TADA-Ed للتنقيب والعرض، Sequential Mining tool لتنقيب الأنماط، MINEL لتنقيب طرق التعلم، CIECoF لتنقيب علاقات الارتباط، Simulog لكشف أنماط السلوك غير المتوقع.

# الفصل الثاني: الدراسة المرجعية.

## مقدمة:

يتناول هذا الفصل دراسة لبعض الأنظمة المشابهة التي عملت على التنقيب في المعطيات التعليمية. يجدر الذكر أن عملية التنقيب تختلف باختلاف المجال الذي يتم التنقيب فيه بالإضافة لطبيعة البيانات وعديد من العوامل الأخرى. إذ يصعب تعميم تجربة شملت نظام معين على باقي الأنظمة، وإنما يمكن الاستفادة منها بمنهجية العمل المتبعة والنتائج المرجو الوصول إليها.

تتصف معظم أنظمة التحليل والتنقيب الموجود حالياً لاحتياجها أشخاص ذوي خبرة للتعامل مع هذه الأنظمة نتيجة تعقيد هذه الأنظمة، إذ أن استخدام المدرسين مثلاً لمثل هذه الأنظمة دون الخضوع لتدريب خاص عملية شبه مستحيلة، بالإضافة إلى أنها تستهلك وقت كبير جداً. علاوة على ذلك، يمكن للمدرسين وحتى زملاء الطلاب في بعض الأحيان الاطلاع على معلومات شخصية غير ضرورية وتجاهل قضايا خصوصية البيانات في تصميم معظم بيئات التعلم الافتراضية. وهنا يجب الانتباه إلى الجانب القانوني، فعلى سبيل المثال في ألمانيا، لا يحق للمدرسين الوصول إلى *جميع* البيانات الناتجة عن أفعال الطالب خلال ارتياد دورات أونلاين، وإنما يمكنهم الوصول فقط إلى البيانات ذات الصلة بالتدريس بشكل يؤمن الشفافية للطلاب.

إلا أن عملية التحليل تهتم بكسب معرفة واسعة حول بيانات الطلاب، لذلك ولضمان خصوصية الطلاب، يتم اللجوء إلى تزييف هوية الطلاب قبل البدء بالمعالجة، على سبيل المثال، يتم استخدام التجزئة hash بدلاً من الهويةID ، وتقديم النتائج على شكل تصورات التي تظهر تشمل المجموعات وعدم التركيز على طالب واحد.

## بعض الأنظمة المشابهة:

### [Open University Analyse:](https://analyse.kmi.open.ac.uk/)

هدف المشروع: يهدف المشروع للكشف المبكر عن الطلاب المعرضين لخطر الفشل والرسوب. احتمالات فشل أو تقصير جميع الطلاب تكون متاحة أسبوعياً لمدرسي المواد وفِرق دعم الطلاب لتقديم الدعم المناسب لهؤلاء الطلاب.

البيئة الحاضنة: تم دراسة المشروع على ضوء الجامعة الافتراضية Open university الواقعة في المملكة المتحدة.

عن المشروع: أُجري عدد كبير من الأبحاث في سبيل إنجاز المشروع، ومر المشروع بعدة أطوار، نذكر فيما يلي أهم سمات كل مرحلة:

* الطور عام 2013

كان تركيز هذه المرحلة على انعدام التواصل المباشر وجهاً لوجه في الجامعة الافتراضية OU الذي يؤدي لخلق صعوبات بمعرفة المشاكل والعقبات التي يواجهها الطلاب والتي تؤدي لفشلهم. بالإضافة لأعداد الطلاب الكبيرة وضيق وقت المعلمين وعدم إمكانهم ملاحقة حال جميع الطلاب. بالتنبؤ بالطلاب الذين هم بخطر الفشل يتم تقديم الدعم الكافي لهم لتجنب هذا الفشل. نتيجة لاحتواء البيئة التعليمية الافتراضية على معلومات كثيرة شخصية عن الطالب وسلوكه التعليمي ونتائج الوظائف والامتحانات وغيرها، طورت الجامعة الافتراضية OU نماذج توقع تأخذ بعين الاعتبار نوعين من البيانات: أداء الطالب بالامتحانات والمذاكرات الدورية، وتفاعله مع بيئة التعلم الافتراضية. تم استبعاد البيانات المتعلقة بالطلاب قليلي التفاعل مع بيئة التعلم الافتراضية، لأنه وبعد الأبحاث وُجد أن بعض الطلاب لا تُظهر تفاعل كبير مع بيئة التعلم بالرغم من النتائج الجيدة التي يحققوها في المقرر نتيجة لاستخدامهم مراجع أخرى offline دون استخدام بيئات التعلم هذه أو نتيجة كون الطالب يعيد المقرر نتيجة رسوبه به سابقاً. لحل هذه المشكلة تم اتباع طريقة تحتفظ بالسلوك العام للطالب باستخدام learning profile إذ أن سلوك الطالب بشكل عام يكون ثابت "من أجل مقرر ما، ويتغير هذا السلوك من مقرر لآخر"، لتجنب مقارنة سلوك الطالب مع زملائه نتيجة لاختلاف السلوك من فرد لآخر (فمن الممكن أن يحصل طالبان على نفس العلامة بالرغم من الاختلاف بينهما بكمية التفاعل مع بيئة التعلم) يتم مقارنة الفرد مع سلوكه السابق لاكتشاف تغير سلوك الطالب نتيجة لصعوبات يواجهها. كما أنه يمكن الاستفادة من تفاعل *المدرسين* أيضاً في تحليل وتخمين نتائج الطلاب.

* الطور عام 2014

نتيجة الأبحاث تبين أن الطالب الذي يفشل بأول تقييم له غالباً يكمل بالفشل بكامل المقرر، لذلك كان لا بد من اتخاذ إجراءات لمنع فشل الطلاب في أول تقييم لهم وذلك بالتنبؤ بحال الطلاب قبل أول تقييم. تم الاعتماد في عملية التنبؤ هذه على بيانات ديموغرافية شخصية، والبيانات المخزنة في بيئة التعلم الافتراضية وتم الاعتماد على عدة خوارزميات machine learning مثل k Nearest Neighbours, Classification and Regression Tree, Bayes network.

أظهرت الأبحاث أن نسبة الطلاب الذين يكملون في الدورات على شبكة الإنترنت (كالدورات على منصات MOOC) لا تتجاوز الـ 5%، إلا أن هذه الحالة تكون أفضل بكثير في الجامعات التقليدية والـ OU. عادة ما يتم تقسيم الطلاب لمجموعات صغيرة (عشرات الطلاب) حسب موقعهم الجغرافي، وتعيين مشرف يساعدهم ويجيب على أسئلتهم، إلا أن المشكلة الحقيقية تكمن بمساعدة الطلاب الذين بحالة خطرة واكتشاف هذا الشيء بمراحل مبكرة، خاصة وأن كل مقرر يحوي آلاف الطلاب، فمن الصعب توفير موارد كافية لكل هذه أعداد، وتحديد الأشخاص الذين بحالة حرجة للقيام بالتدخل بالوقت المناسب.

نتيجة لكمية البيانات الهائلة الناتجة عن بيئة التعلم الافتراضية تم غربلة هذه البيانات، إذ تم احتساب:

* عدد النقرات في الأسبوع.
* عدد النقرات بالأسبوع وحسب المحتوى الذي تم الضغط عليه.
* Flag يدل عن كون الطالب نشيط وكثير التفاعل في بيئة التعلم أم لا.

وتم تصنيف المحتوى وفق 11 نوع (اختبارات، ويكي، مراجع..) وتحديد نسبة مساهمة التقصير بكل محتوى بفشل الطالب، واعتمدوا على Bayes Theorem لتحديد احتمالية فشل الطالب. أما عن تحليل سلوك الطالب ببيئة التعلم الافتراضية فاستخدموا General Unary Hypotheses Automaton “GUHA” && modelling based on Markov chains. السيئة التي كانت بالـ GUHA هي خسارة المعلومات التي لها علاقة ببعد الزمن، لذلك تم العمل على نموذج آخر يقوم على التحليل بالاعتماد على سلاسل ماركوف.

* الطور عام 2015

اعتمد البحث على نوعين من البيانات: بيانات ديموغرافية وبيانات ناتجة عن تفاعل الطلاب مع بيئة التعليم الافتراضية، باستخدام Bayesian approach يتم استخراج أهم الأنشطة من التفاعلات بحيث يتم استخدام هذه الأنشطة مع البيانات الديموغرافية لبناء أربعة نماذج تنبؤيه:

* + Bayesian classifier
  + Classification and regression tree (CART)
  + k Nearest Neighbours (k-NN) with demographic/static data
  + k-NN with VLE data

يتم استخدام البيانات الديموغرافية (العمر، التعليم السابق، عدد المواد التي قام الطالب بالتسجيل بها...) للقيام بعملية التنبؤ في مرحلة لم يقم الطالب فيها بالتفاعل مع النظام ولم يتم القيام بأي تقييمات بعد، مع مرور الوقت وعندما يبدأ الطالب بالتفاعل مع البيئة الافتراضية يتم تقليل وزن المعطيات الديموغرافية ويصبح التركيز الأكبر على التفاعل مع بيئة التعلم الافتراضية. تكمن الصعوبة بإعطاء **الأوزان** للتفاعلات بين الطالب والنظام التعليمي وكيفية معرفة أيهما أكثر أهمية. صعوبة أخرى تكمن في عملية تحضير البيانات التي تتم بطريقة يدوية. والعمل على رفع معدل الطالب وعدم الاكتفاء بالتنبؤ بنجاح أو رسوب الطالب.

* الطور عام 2016

شمل التطبيق الذي تم تقديمه باستخدام لغة R التحليلية 32593 طالب تشمل بيانات عن تقييمات الطلاب وتفاعلهم مع البيئة التعليمية الافتراضية، وقد تم إخفاء هوية الطلاب باستخدام أداة إخفاء هوية البيانات ARX. جرى في ذلك العام إطلاق Open University Learning Analytics Dataset OULAD وإرفاق شرح عن الجداول والبيانات وكيفية المعالجة.

يشكل OULUD *جزء* من البيانات المجنية من الجامعة الافتراضية OU، إذ تحوي معلومات ديموغرافية ومعلومات عن أداء الطلاب خلال المقررات وتفاعلهم مع بيئة التعلم. تقدم هذه البيانات معلومات عن أداء الطلاب وتقدم فرصة لخلق جيل جديد من نظم إدارة التعلم. تتركز قاعدة المعطيات هذه حول الطالب بدل من المقرر. علاوة على ذلك فإن الجدول أو الملف الرئيسي بقاعدة المعطيات هي معلومات الطالب.

### [Completing the Loop:](http://melbourne-cshe.unimelb.edu.au/research/edutech/completing-the-loop)

هدف المشروع: يعمل المشروع على **تحليل** عملية التعلم في بيئات التعلم الافتراضية وتقديم نتائج هذا التحليل **بالطرق الأمثل** بهدف الوصول إلى سُبل وإجراءات تفيد المعلمين والمدرسين في رفع سوية التعليم.

البيئة الحاضنة: تم تنفيذ المشروع على أربع مقررات ضمن ثلاث جامعات: مقررين في جامعة ملبورن Melbourne، مقرر في جامعة ماكواري Macquarie، وآخر في جامعة جنوب أستراليا the University of South Australia .

عن المشروع:

تم التعاون في المشروع مع اثني عشر مدرس من ثلاث جامعات، امتد من بداية العام الدراسي 2014 إلى منصف العام الدراسي 2016. نوعية البيانات المقدمة من الجامعات كانت مختلفة لاختلاف طرق التعليم في الجامعات الثلاث. قدم بعض المشاركين اقتراح مصادر إضافية للبيانات غير نظم إدارة التعلم، كالبيانات الناتجة عن أنظمة التقييم، معلومات شخصية عن الطلاب مستقاة من نظم المعلومات الخاصة بالجامعة student information system، ومعلومات الحضور. من وجه نظر المعلمين كانت المشكلة الأساسية هي حاجة المعلمين للتدرَب على استخدام نتائج التحليل وفهمها نتيجة لكون نتائج التحليل غير مفهومة لغير المختصين، إضافة للوقت المطلوب لجمع وتحليل البيانات واتخاذ القرار بناء على نتائج عملية التحليل. أما عن الأداة المبنية فكانت باستخدام Python, Django, Pandas, MySQL للوصول إلى خدمة مرتكزة على الويب، تستخدم بيانات مستوردة من Blackboard وMoodle.

عادة ما يكون هناك فجوة معرفية في أنظمة التحليل ناتجة عن عدم توافر جسر يصل بين المعلومات الناتجة عن تحليل التعلم ونوع النظام والعمليات التعليمية المتبعة في المؤسسات التعليمية، لذلك يهتم المشروع بشقين أساسيين: عملية *تحليل التعلم Learning Analytics* التي تسعى لتحسين العملية التعليمية عن طريق كشف المشاكل الناتجة عن التعلم الإلكتروني وتقديم تغذية راجعة Feedback لكل من الطالب والمعلم، وعملية *تصميم التعلم Learning Design* التي تشمل التخطيط لكيفية سير العملية التعليمية والأنشطة التي تتضمنها. يجب المكاملة بين العمليتين السابقين للوصول إلى فهم أعمق لحالة الطلب، *فسلوك* الطالب في الأنظمة التعليمية لا يعبر بالضرورة عن حالته، على سبيل المثال: تحميل محاضرة لا يعني أن الطلب قد قرء وفهم المحاضرة، وبالتالي لا يجب الاعتماد بشكل كامل على نتائج التحليل بل يجب توخي الحذر عند القيام بالتحليل، والاستعانة بالـ *Learning Design* التي من شأنها أن تساعد في تحديد المعايير التي تجري من خلالها تقييم عملية التعلم.

##### Conversational Framework Laurillard’s

تم استيحاء فكرة المشروع من Laurillard’s conversational framework والتي ترى أن عمليات التعلم لدى الطلاب بحاجة لأن تكون مدعومة بحلقة متكررة من التفاعل والحوار وتبادل المعلومات بين المعلمين والطلاب. يبدأ تفاعل التعليم عند تصميم النشاط التعليمي وتقديمه من قبل المعلمين إلى الطلاب. تلي هذه العملية انخراط الطلاب بهذا النشاط (كقراءة مصادر التعلم، المشاركة في المناقشة، إلخ) باستخدام فهمهم الحالي للموضوع. تقدم طرق انخراط الطلاب بالنشاط التعليمي شكلاً من أشكال التغذية الراجعة للمعلم. بناء على هذه التغذية الراجعة يمكن للمعلم إعادة تقديم النشاط، وتقديم بعض أشكال المعالجة وبذلك تبدأ حلقة جديدة.

هيكلية المشروع: يحوي المشروع مستويين من المستخدمين:

* مدير: يملك صلاحيات
* إضافة مستخدم جديد
* إضافة مجموعة مستخدمين يملكون أذونات وصول محددة.
* منح حق الوصول لمواد محددة
* إنشاء مقرر
* إنشاء حدث لمقرر، وتتراوح أنواع الأحداث بين أحداث تتكرر أسبوعياً (كمحاضرة كل يوم ثلاثاء)، أحداث خلال الفصل (كرحلة ميدانية في الأسبوع الرابع)، وأحداث تقديم (كالاختبارات الممتدة من الأسبوع السابع حتى التاسع من الفصل).
* مدرس: يمكنه الوصول إلى أداة الدعم التربوي وأداة تحليل التعلم من أجل جميع المواد التي يملك صلاحيات الوصول لها.

يقدم المشروع أداتين:

* Pedagogical Helper Tool: توفر للمعلمين فضاء لتبيان العلاقة بين أهداف التعلم، طرائق التعلم، والتكنولوجيا المستخدمة في المقرر. يقوم المدرس بإضافة أهداف التعلم (يمكن إضافة أكثر من هدف) ومن أجل كل هدف يتم إضافة النشاطات التي سيتم اتباعها لمساعدة الطلاب للوصول للهدف (كتحضير المحاضرة قبل ارتيادها). يتم بعد ذلك إضافة مصادر للتعلم من أجل كل نشاط. وبذلك يتم تمثل مرئي للعلاقات ما بين *تصميم* التعلم (الممثل بأهداف ونشاطات التعلم) *وتحليل* التعلم (الممثل بمصادر التعلم). تكون مخرجات هذه المرحلة كخريطة تساعد المعلمين بتفسير البيانات الناتجة عن عملية التحليل.
* Learning Analytics Tool: صُممت لعرض البيانات الناتجة عن التحليل للمعلمين بطريقة مفهومة. تقسم هذه الأداة لثلاث أقسام رئيسية: *course dashboard* تقدم ملخص لتفاعل جميع الطلاب مع نظام إدارة التعلم خلال أسابيع محددة أو خلال كامل الفصل، *course access* تشمل معلومات عن المحتوى، التفاعل والتقدير، *وstudents* التي تسمح بالوصول إلى طالب محدد واستعراض تفاعله مع نظام إدارة التعلم خلال مدة المقرر.

# الفصل الثالث: الدراسة التحليلية.

## أنواع البيئات الافتراضية التعليمية:

فيما يلي ذكر لأهم البيئات الافتراضية التعليمية:

### Massive Open Online Courses (MOOC):

إن الدورات الافتراضية على شبكة الإنترنت في تزايد مستمر، كما أن الاهتمام بها من قبل المجتمع العلمي يزداد أيضاً. يمكن تعريف الـ MOOC بأنها دورات على شبكة الإنترنت تهدف للتفاعل الكبير والواسع بين مرتادي هذه الدورات. إن إتاحة هذه الدورات على شبكة الإنترنت يجعل من الممكن وصول أي فرد يملك اتصال بشبكة الإنترنت التسجيل بهذه الدورات مجاناً. بعض الأمثلة عنها Udacity, MITx, EdX, Coursera, Udemy. إن هذه المنصات تشابه لحد كبير استخدام نظم إدارة التعلم LMS ولكن مع زيادة عدد المستخدمين بشكل كبير، وبذلك تكون كمية البيانات الناتجة عن هذه الأنظمة ضخمة جداً مما يجعل الحاجة لاستخدام التنقيب في البيانات والتحليل ماسة لمعالجة هذه البيانات.

### Intelligent Tutoring Systems:

أنظمة الإرشاد الذكية هي أنظمة تقدم تعليمات وملاحظات **مباشرة** للطلاب (عادة ما تكون دون تدخل بشري من المعلم). ويهدف البرنامج إلى توفير فرص الحصول على تعليم عالي الجودة إلى كل طالب، وبالتالي إصلاح النظام التعليمي برمته. تتغير نماذج التفاعل بناءً على النماذج الفردية. يؤدي قدرة هذه الأنظمة على تسجيل وتجميع تفاصيل التفاعلات مع أعداد كبيرة من الطلاب إلى خلق مجموعة ضخمة من البيانات. على الرغم من أن هذه الأنظمة تسجل كل تفاعلات الطلاب في ملفات أو قواعد بيانات، هناك مخازن بيانات أخرى متاحة في هذه الأنظمة، على سبيل المثال، نموذج المجال الذي يشتمل على مجموعة من القيود ذات الصلة بنطاق المدرس، ومجموعة البيانات التربوية التي تحتوي على مجموعة من المشاكل وإجاباتها، ونموذج الطالب الذي يقوم بتخزين المعلومات حول كل طالب مع مراعات جميع القيود. هناك أبحاث جارية لتحسين فعالية هذه الأنظمة.

### Adaptive and Intelligent Hypermedia Systems:

أنظمة الوسائط الفائقة المتكيفة والذكية توفر بديل لمنهج just-put-it-on-the-Web التقليدي في تطوير المناهج التعليمية التربوية. تحاول هذه الأنظمة أن تكون أكثر قابلية للتكيف من خلال بناء نموذج للأهداف، والأفضليات، ومعرفة كل طالب على حدة واستخدام هذا النموذج في التفاعل مع الطلاب من أجل التكيف مع احتياجات هؤلاء الطلاب. تكون البيانات القادمة من هذه الأنظمة غنية ويمكن أن تؤدي إلى تحليل أفضل من التحليل القائم على البيانات الناتجة عن أنظمة التعليم المعتمدة على الويب التقليدية. إن البيانات المتاحة في هذه الأنظمة مشابهة لبيانات أنظمة الإرشاد الذكية ITS.

### Test and Quiz Systems:

أنظمة الاختبار هي من أكثر الأدوات انتشار واستخدام في مجال التعليم كما أنها يتم تطويرها بشكل فعال وجيد. يتكون الاختبار من سلسلة من الأسئلة بغرض جمع المعلومات من الممتَحنين. الهدف الرئيسي من هذه الأنظمة هو قياس المستوى المعرفي للطلاب بما له علاقة بموضوع أو أكثر. هناك أنواع مختلفة من الأسئلة مثل الإجابة بـ نعم/لا، أسئلة الاختيار من متعدد MCQ، ملأ الفراغات، الأسئلة الكتابية وغيرها. وتخزّن نظم الاختبار كمية كبيرة من البيانات، مثل الأسئلة، أجوبة الطلاب، نتائج الطلاب، وبعض الإحصائيات.

### Learning Management Systems:

هي نوع خاص من أنظمة التعلم عبر الإنترنت تتيح إدارة، توثيق وتتبع البرامج التعليمية، الفصول الدراسية، الأحداث عبر الإنترنت والمحتوى التعليمي. كما أنها توفر مجموعة كبيرة ومتنوعة من القنوات ومساحات العمل لتسهيل تبادل المعلومات والتواصل بين جميع المشاركين في العملية التعليمية.

بعض الأمثلة عن نظم إدارة التعلم التجارية: Blackboard و Virtual-U في حين أن هناك أنظمة أخرى مجانية مثل Moodle, Ilias, Sakai and Claroline.

تُراكم هذه الأنظمة بيانات هائلة ناتجة عن تسجيل كافة أنشطة وعمليات تفاعل الطلاب مع الأنظمة التعليمية، وعادة ما يوجد أدوات تتبع مضمّنة في هذه الأنظمة تسمح للمدرس بعرض بيانات إحصائية للبيانات. يمكن أن يتم تسجيل أي نشاط يتم في النظام مثل القراءة، الكتابة، الاختبارات، والتعليق على الأحداث. كما أن أنظمة التعلم هذه توفر قاعدة بيانات علائقية تخرن فيها معلومات الطالب في جداول مختلفة كمعلومات المستخدم الشخصية (الملف الشخصي)، النتائج الأكاديمية (الدرجات)، وبيانات تفاعل المستخدم (التقارير).

نتيجة لاحتضان هذه الأنظمة بيانات ضخمة التي يتم توليدها يومياً من بيانات شخصية عن الطلاب ونتائج الطلاب الأكاديمية وبيانات التفاعل، يصعب إدارة هذه البيانات بشكل يدوي. وبالرغم من أن بعض أنظمة إدارة المحتوى توفر أدوات إعداد تقارير إلا أنها لا تسمح للمعلمين بتتبع وتقييم بدقة جميع الأنشطة التي يقوم بها المتعلمون وتقييم هيكلية ومحتوى وفعالية المادة العلمية، كما أنه يصعب على المعلمين استخلاص معلومات مفيدة من هذه الأدوات مع وجود عدد كبير من الطلاب، وهنا تبرز أهمية التنقيب.

* **Moodle**: Modular Object Oriented Developmental Learning Environment

نظام إدارة تعلم مفتوح المصدر، يساعد المدرسين على خلق بيئة تعليمية تفاعلية فعالة على شبكة الإنترنت. يدعم حوالي 75 لغة فيما يزيد عن 160 دولة. ويعتمد في تصميمه على نمط تعلّم يدعى social constructionist pedagogy الذي يقوم على أن أفضل العمليات التعليمية تكون عند التفاعل مع: المحتوى التعليمي والأقران حول المحتوى التعليمي.

يحوي مودل على:

-خمس أصناف للأدوات التعليمية: الصفحات النصية، صفحات الويب، رابط لشيء آخر على الويب، view into one of the course’s directories وعرض لنص أو صورة.

-ست طرق للتفاعل مع المقرر: الوظائف، الخيارات، الصحف، الدروس، الاختبارات والدراسات الاستقصائية.

-خمس طرق للتفاعل مع الآخرين: الدردشة، المنتدى، glossary، الويكي وورشات العمل.

يخزن مودل كل تحركات الطلاب وكل النقرات التي يقومون بها في النظام، ويمكن فلترة هذه المعلومات على حسب المقرر، النشاط، الطالب، واليوم. كما يمكن للمعلمين استخدام هذه البيانات لمعرفة الطلاب الأكثر نشاط في المقرر. لا يتم تخرين البيانات السابقة على شكل نصوص وإنما يتم تخزينها في قاعدة معطيات علائقية. أفضل النظم إدارة قواعد المعطيات الداعمة لمودل MySQL, PostgreSQL إلا أنه يمكن استخدام نظم أخرى مثل Oracle, Access, Interbase. يحوي مودل حوالي 145 جدول مترابط، إلا أن عملية التنقيب لا تحتاج إلى كل هذه البيانات وهنا يبرز دور مرحلة ما قبل التنقيب. نتيجة لبنية مودل تكون كمية الجهد المبذول في مرحلة ما قبل المعالجة أقل من الأنظمة التعليمية الأخرى. إلا أن بعض العمليات لا تزال مطلوبة مثل اختيار البيانات التي سيتم عملية التنقيب عليها وتحويل البيانات الرقمية إلى وصفية ونقل وتحويل البيانات للصيغة الأمثل لعميات التنقيب كصيغة ARFF(Attribute-Relation File Format) مثلاً.

تطبيق تقنيات التنقيب على بيانات Moodle:

* الإحصائيات: لا يؤمن مودل إحصائيات تزود المدرسين بتقارير محددة عن تفاصيل إحصائية لأداء كل طالب. وإنما يزود بإحصائيات لها علاقة فقط بأداء الطلاب (كالعلامات والاختبارات)، فيمكن للمدرسة معرفة إداء الطلاب في كل سؤال في الاختبار، مما يساعد المعلمين على تحديد الأسئلة الواجب تعديلها أو حذفها نتيجة لصعوبتها أو سهولتها الفائقة.
* العرض: يتعلق برسومات الحاسوب وواجهات المستخدم ويهتم بطرق عرض البيانات بشكل رسومي أو متحرك ليسهل على المستخدم فهمها. تبرز فعالية هذه الطريقة عند التعامل مع كمية كبيرة من البيانات ولتمثيل البيانات متعددة الأبعاد. لا يؤمن مودل طرق مرئية لعرض هذه البيانات غير النصوص، إلا أنه يمكن استخدام بعض الأدوات مثل GISMO الذي يمثل البيانات الناتجة عن تتبع الطلاب في مودل بطرق متعددة مرئية ومفهومة.
* العنقدة: تتصف بأنها [*unsupervised learning*](http://en.wikipedia.org/wiki/Unsupervised_learning)، تعمل العنقدة على كشف العلاقات بين العناصر بهدف تجميع العناصر المتشابهة. المبدأ المتبع في العنقدة هو زيادة التشابه في العنقود الواحد وتقليص التشابه بين العناقيد المختلفة. في التعلم الافتراضي يمكن الاستفادة من العنقدة في إيجاد الطلاب ذوي الخصائص التعليمية المشتركة وتشجيع التعلم التعاوني. يمكن بعد تجميع الطلاب في مجموعات تحديد مرشد خاص بهم بناء على الخصائص التي يتمتع بها طلاب المجموعة ومهاراتهم. كما يمكن توجيه اختبارات خاصة بهم. تؤمن الأداة Weka عدة خوارزميات عنقدة. يمكن على سبيل المثال تجميع الطلاب بناء على نشاطاتهم في مودل ونتائج الاختبارات النهائية.
* التصنيف: يتصف بأنه[supervised learning](http://en.wikipedia.org/wiki/Supervised_learning)، يعتمد على معرفة انتماء عناصر جديدة لفئات موجودة ومحددة مسبقاً. يمكن الاستفادة من التصنيف في التعلم الافتراضي في كشف الطلاب ذوي الخصائص المتشابهة ورداة الفعل المتشابهة على الاستراتيجية التربوية المتبعة، وللتنبؤ بأداء الطلاب ونتيجتهم النهائية ولتحديد المتعلمين الذين لا يملكون الدوافع والمحفزات اتخاذ الإجراءات لتخفيض معدلات التسرب. تؤمن الأداة Keel عدة خوارزميات تصنيف. يمكن للمعلم استخدام المعرفة الناتجة عن التصنيف باتخاذ قرارات لها علاقة بالنشاطات المتعلقة بالمقرر ضمن مودل.
* علاقات الارتباط: تعتمد علاقات الارتباط على مدى ارتباط حدوث العنصر x بعنصر آخر y. يمكن الاستفادة من علاقات الارتباط بتحديد المواد التعليمية الأنسب لاقتراحها على المستخدم، ولاكتشاف العلاقات المثيرة للاهتمام من استخدام الطالب للمعلومات بهدف توفير التغذية الراجعة لمدرس المقرر، ولكشف العلاقات بين سلوكيات التعلم. كما يمكن استخدامها لكشف الأخطاء التي يشترك الطلاب بارتكابها. تؤمن الأداة Weka عدة خوارزميات لكشف علاقات الارتباط.
* تنقيب الأنماط التتابعية
* تنقيب النصوص

### أنواع أخرى للأنظمة التعليمية:

إضافة لما سبق يوجد أنواع أخرى من الأنظمة التعليمية مثل بيئات الألعاب التعليمية، بيئات الواقع الافتراضي، مستودعات كائنات التعلم، الويكي، المنتديات، المدونات وغيرها.

## أنواع البيانات:

* البيانات العلائقية
* بيانات المناقلات
* بيانات متسلسلة وذات طابع زمني
* بيانات نصية
* بيانات متعددة الوسائط
* بيانات شبكة الإنترنت

## عملية ما قبل المعالجة:

كلما كانت عملية تحضير البيانات الخام قبل البدء بمعالجتها أكثر، كلما أعطت عملية المعالجة نتائج أفضل. ولكن عملية التحضير هذه تتطلب جهد يدوي كبير، إذ أن هذه العملية تستهلك حوالي 60-90% من الوقت والجهد والموارد. فيما يتعلق بالبيانات الناتجة عن الأنظمة التعليمية نجد أن كمية البيانات ضخمة جداً وربما لا تكون بالشكل الأنسب وبالتالي تكون عملية التحليل غير مجدية. لذلك يجب العمل على تهيئة هذه البيانات للحصول على النتائج الأمثل من عمليات التحليل. هناك أمور أخرى يجب الاهتمام بها كعملية المكاملة بين البيانات الناتجة عن أنظمة مختلفة.

يمكن تلخيص أهم المهام التي يتم إنجازها في عملية ما قبل المعالجة بما يلي:

data cleaning, user identification, session identification, path completion, transaction identification, data transformation and enrichment, data integration, data reduction.