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 التعليمي 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 و معلمين 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.

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 الأخرى الـ online في العمليات التعليمية، أتيحت كمية هائلة من الداتا تبعاً 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:

بالعودة إلى جذور التنقيب عن المعرفة في الأنظمة التعلمية نجد أن الأبحاث في هذا المجال جديدة نسبياً. فبالرغم من أن عمليات البحث والتحليل للداتا الناتجة عن البرمجيات التعلمية بدأت من مدة طويلة، إلا أن الـ 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).

## معنى الـ EDM:

يوجد عدة تعريفات لها، إلا أن تعريفها حسب journal of educational data mining and Baker: هو 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 التعليمة والبيئة التي تحضنها.

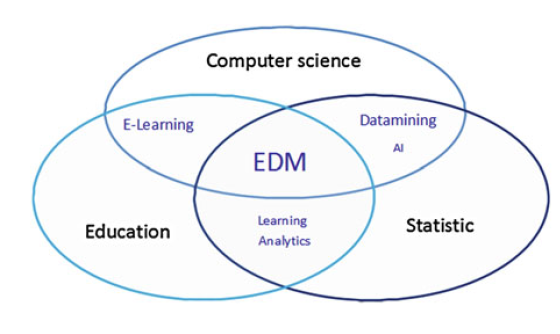
مع عدم وجود جدار فصل حقيقي بين المجالين ووجود العديد من الأهداف المشتركة بينهما تزداد صعوبة الفصل وتميز كون تطبيق يتبع لأحد المجالين أو للآخر.

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

كمجال متعدد الاختصاصات يستخدم الـ EDM طرق وتقنيات مختلفة مثل:

statistics, machine learning, data mining, information retrieval, recommender systems, psycho-pedagogy, cognitive psychology, psychometrics, etc

وعملية اختيار التقنيات المستخدمة تعتمد على القضية التعليمية المطروحة.



## 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
* توقع مستوى الطلاب التعليمي ونتائجهم
* إيجاد الاقتراحات للطلاب
* تحليل سلوك المتعلم
* التواصل مع الـ stockholders
* تحليل هيكلية المجال
* تحسين ومتابعة المسارات التعليمية
* دراسة التأثيرات لمختلف pedagogical support الذي يتم تقديمه من قبل نظم التعليم
* دفع المعرفة العلمية حول التعليم والطلاب من خلال بناء واكتشاف وتحسين نماذج الطالب والمجال وال pedagogical support

من الجدير بالذكر أن أهداف تنقيب البيانات التعليمية سابقة الذكر تهدف إلى تحسين النظم التعليمية بشكل عام وتحسين النظم التعليمية المعتمدة على الحاسوب بشكل خاص لذلك نمذجة الطالب هي المفتاح الأساسي لتحقيق العديد من أهداف ومهام تنقيب البيانات التعليمية.

## الطرق المستخدمة في تنقيب البيانات التعليمية:

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

* **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). ومثال على طرق التوقع هي Classification , Regression , Density Estimation ومثال 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. في تنقيب البيانات تستخدم هذه الطريقة 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 و casual data mining. في تنقيب البيانات التعليمية يستخدم تنقيب العلاقات لتحديد العلاقات بين نشاطات الطلاب على الشبكة وعلاماتهم مثلاً، أو model learners' problem solving activity sequences.
* **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، والقواعد 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 التقيب في النصوص لتحليل مضمون المناقشات، المنتديات، الدردشة، 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.