علم بيانات التعليم: الخلفية والتطور

أ.د. محمد محمد الهادي
أكاديمية السادات للعلوم الإدارية

المستخلص

هذه المقالة توضح أبعاد مجال ناشئ حديثا أطلق عليه «علم بيانات التعليم Education Data Science»، وهو مجال متعدد تخصصات الاستفسار ويجمع معا كل من علم الحاسب الآلي، علم البيانات، تحليلات البيانات الضخمة، تحليلات التعلم، الذكاء الاصطناعي، الإحصاء والعلوم الاجتماعية الأخرى لفحص وفهم الظواهر الاجتماعية والفنية. ويستخدم الباحثون والممارسون مجموعات مختلفة من الإجراءات والأساليب لجمع، تنظيم، تداول، وتفسير مصادر البيانات التعليمية الثرية. إلي جانب ذلك يعرض علم بيانات التعليم أساليب دمج البيانات الضخمة والمتنوعة معا، ضان اتساق مجموعات البيانات، وخلق التصور الموحد للمساعدة في فهم البيانات المعقدة. علاوة على ذلك، في هذا العلم الناشئ يتبني علمائه استخدام النماذج الرياضة لتوصيل الأفكار/النتائج لعلماء وأخصائي التعليم الآخرين، وإذا لزم الأمر لأصحاب المصالحة غير الخبراء. ويعتبر هذا العلم أحد المجالات الفرعية لعلم البيانات Data Science. ويعتبر علم بيانات التعليم مجال بحثي مهتم بتطوير الطرق لاستكشاف البيانات التعليمية على نطاق واسع بشكل متزايد بغية فهم الطلاب والأوضاع التي يتعلمون فيها بشكل أفضل. وبذلك، ينظر لعلم بيانات التعليم بمعني ضيق بأنه يبحث في الأساليب الحوسبية والاحصائية عن الظواهر والأسئلة التربوية، كما ينظر إليه أيضا بمعني أشمل كمظلة لكم من الأساليب الحوسبية التي تستخدم لتحديد أشكال البيانات، المقاييس، التنبؤات، والتجارب الجديدة في التعليم التي ليست فقط أسئلة بحثية قديمة يجري تحليلها بطرق جديدة، لكن أيضا الأسئلة الناشئة الجديدة على أساس بيانات واكتشافات جديدة من أساليب علم بيانات التعليم. وتحدد هذه النظرة العامة مجال علم بيانات التعليم. وقد غير علم بيانات التعليم الطريقة التي يتفاعل فيها الطلاب مع المدرسين وتؤدي لتقييم أدائهم. وصار المدربون والمعلمون يستخدمون علم بينات التعليم من أجل تحسين طرق تدريسهم. وبذلك يساهم علم بيانات التعليم لإنشاء نمذجة تنبؤية التي يمكنها توقع معدل تسرب الطلاب بناء على أدائهم مع إبلاغ معلميهم لاتخاذ الاحتياطات اللازمة.

الكلمات الرئيسية: علم بيانات التعليم، علم البيانات، تحليل الشبكات، معالجة اللغة الطبيعية، تحليلات التعلم، التعلم الآلي، التنقيب عن البيانات التعليمية.

١. خلفية علم بيانات التعليم

التعليم هو مفتاح تشكيل حياة البشر، حيث أنه لده القدرة على تغيير وإثراء حياة الناس. ومنذ فجر الحضارة البشرية تطور البشر من خلال التعليم، وطوروا آليات لتحسين التعليم. وفي القرن الحادي والعشرين الحالي، حيث تتواجد البيانات بكثرة في كل مناحي الحياة، وليس

التعليم استثناء من ذلك. ومع التقدم في أساليب الحوسبة، صار ممكنا استيعاب كل المعلومات من خلال منصات البيانات الضخمة القوية. وفي هذا السياق تم استخدام علم البيانات مشكلا علم بيانات التعليم لكي تؤدي دورا رئيسيا في جعل التعليم أحسن للمجتمع.

وحيث أن علم بيانات التعليم منبثق من علم البيانات الذي يعزي بدايته في وقت مبكر لستينيات القرن العشرين الماضي. وأول خطوة في هذا الاتجاه كانت في إطار مستقبل تحليل البيانات والمناقشات التي أبديت في سبعينيات وثمانيات القرن العشرين عن هذا العلم كاسم مستعار لعلوم الحاسب الالي والاحصاء على التوالي (Donoho, 2017). وقد أدت نشأت ظهور علم البيانات في العقد الثاني من القرن الحادي والعشرين كأداة منهجية متعددة التخصصات ترتبط بالبيانات الضخمة وتكنولوجياتها، التي تم استخدامها لبيانات التعليم الضخمة وتحليلات التعلم، البارزة في علم البيانات من خلال بحوث التعليم لدراسة في مجموعة نشر مركز معلومات موارد التعليم «إيريك ERIC». وفي إطار مطبوعات هذا المركز صار في الإمكان تتبع أبعاد استخدام علم البيانات في التعليم وتحديدها من خلال معدل اعتماد خمس تخصصات بارزة هي: تحليلات التعلم Learning Analytics، التعلم الآلي Machine Learning، الذكاء الاصطناعي AI، علم البيانات Data Science، ومعالجة اللغة الطبيعية Natural Language Processing (NLP). ومعدل تطبيق هذه التخصصات الخمسة يمكن مقارنتها مع النمو الشامل في بحوث التعليم حيث يمكن التعرف على أربع اتجاهات رئيسة هي:

أولا: تعرض المطبوعات المنشورة من مركز إيريك ERIC هذه التخصصات الخمسة في وقت مبكر من عام ٢٠٠٠ حتى عام ٢٠١٠ التي تم استخدامها باعتدال كبير.

ثانيا، بعد عام ٢٠١٠ صار هناك ارتفاع حاد في المطبوعات المنشورة التي تستخدم هذه التخصصات. ومن المحتمل أن هذا التعزيز يعزي إلى الارتفاع السريع في شعبية التعلم الإلكتروني واستخدام أدوات المواد الدراسية المفتوحة على الخط MOOCs بين عامي ٢٠٠٨ و٢٠١٢ (Clow, 2013).

ثالثا: تبين المنحدرات المتزايدة لكل منحني أن معدلات نمو علم بيانات التعليم صارت تتسارع مع مرور الوقت. وترتبط سرعة نمو علم بيانات التعليم بأن المطبوعات الصادرة في التعليم يمكن الحكم عليها في الحقيقة من عدد المقالات التي تشير إلى تحليلات التعلم زادت من ٠,٠١٪ في عام ٢٠١٠ إلي حوالي ٠,٣٥٪ في عام ٢٠٢٠ أي بحوالي ٣٥ ضعفا.

رابعا: الأعداد المطلقة الصغيرة نسبيا للمقالات المنشورة التي تشير لعلم بيانات التعليم لا تزال مجالا فرعيا ناشئا في إطار مجال أكبر خاص بالتعليم.

أي أن هناك نقلة نوعية في الدرجة المتعلقة ببحوث التعليم تستضيف دراسات بيانات ضخمة كثيفة مستوحاة من الابتكارات المنهجية من علم الحاسب الالي وعلم الإحصاء. وعلي ذلك، يمكن الإشارة الي البحوث التي تقود هذا التحول كعلم بيانات التعليم. وصار هذا الموضوع يلتقط العديد من المجالات المترابطة معا للنمو. وأصبح في الإمكان اعتبار أول ظهور لذلك متمثلا في التنقيب عن بيانات التعليم Education Data Mining (EDM) وتحليلات البيانات Data Analytics مع تحليلات التعلم Learning Analytics.

٢. التنقيب عن بيانات التعليم وتحليلات البيانات

نشأ التنقيب عن بيانات التعليم حوالي عام ٢٠٠٥ من خلال العمل حول المعلمين الممكن إدراكهم، والنمذجة التنبؤية التي يمكن أن توافر بيانات دقيقة عن نشاط الطلاب (Piety et al, 2014). مفاهيم طرق تنبؤ التنقيب على البيانات، طرق اكتشاف الهيكل، التنقيب عن العلاقات، نموذج التقطير، وتقطير البيانات للحكم البشري قد تم استيعابها كلها. وفي بعض الأحيان، تكون ممتدة من علم الحاسب الآلي المعاصر وبحوث العلم المعرفي (Cognitive Science Baker, 2013).

وبذلك اعتبر كلا من التنقيب على بيانات التعليم وتحليلات البيانات من أهم الأدوات والأساليب في علم بيانات التعليم، حيث أن كلا من التنقيب على البيانات وتحليلات البيانات تم البحث فيهما من خلال التجريب في المعامل وتم اعتمادهما للصناعات المتنوعة لتوليد ذكاء أعمال من أجل تحسين اتخاذ القرار. وبدأت مؤسسات التعليم العالي خصوصا في استخدام التحليلات لتحسين الخدمات التي تقدم ولرفع درجات الطلاب واستبقائهم في الدراسة بدلا من تسربهم، لذلك يجب تصور طرق استخدام البيانات من نظم التعلم المتاحة على الخط لتحسين عملية التدريس.

ووقد بدأت تجارب التنقيب عن البيانات والتحليلات تتكاثر في التعليم لتعزز الحقائق من ونها مجرد خيال لتحدد إمكانيات البحث والتطبيقات العملية التي ليست سهلة. وحاليا اتجاهات تنقيب البيانات التعليمية لعلم بيانات التعليم تكرس للتركيز على تطوير أدوات جديدة لاكتشاف الأنماط في البيانات. وهذه الأنماط عن المفاهيم الدقيقة المتضمنة في التعليم عموما. ويستخدم التنقيب عن البيانات التعليمية والتحليلات لبحث وبناء النظم في مجالات جديدة التي يمكن أن تؤثر على نظم التعلم المتاحة على الخط. وتعتبر نمذجة المستخدم التي تتضمن ما يعرفه المتعلم، وما هو سلوك ودافعية المتعلم، وما تشبهه خبرة المتعلم، إلى جانب مدي رضا المستخدمين مع التعلم على الخط.

وفي أبسط مستوي، التحليلات تمكن اكتشاف متي يكون الطالب علي الخط بحيث قد يخطئ أو يصيب حتى يمكن تصحيح مساره. والأكثر تعقيدا لدي الطلاب الذين لديهم القدرة لاكتشاف الملل من أنماط النقرات الرئيسية وإعادة توجيه انتباه الطالب. ولأن هذه البيانات جمعت في الوقت الفعلي الحقيقي، توجد إمكانية حقيقية للتحسين المستمر عبر حلقات التغذية الراجعة المتعددة التي تعمل على مستويات مختلفة علي الفور لحل مشكلة الطالب التالية، وللمعلم يوميا للتدريس التالي له، ولأسبوعيا أو شهريا لمدير المدرسة للحكم علي التقدم التعليمي، أو سنويا لمدير المنطقة التعليمية وللدولة ككل لتحسين النظام التعليمي الوطني بشكل جوهري.

٣. تحليلات التعلم

بزغت تحليلات التعلم حديثا مع المؤتمر الذي تعرض لها خصيصا (Simens, 2013)، حيث استعرض يشكل أوسع على الآثار المترتبة على التعلم في العالم الرقمي. وفي الغالب تدرس كيفية نشأت البيانات من النظم الرقمية الممكن استخدامها لوصف أو تسهيل جوانب عملية التعلم أو إلقاء الضوء على البيئات الرقمية (Simens, 2013)، كما أن ظهور برمجيات المواد الدراسية المفتوحة المصدر على الخط MOOCs واستخدام بيانات نظم إدارة التعلم LMS المنشأة من أجل عينات كبيرة من الطلاب حول العالم من مجتمع التحليلات ساهما في ظهور تحليلات التعلم.

وتحليلات التعلم تتناقض مع التنقيب عن البيانات على الأقل في أنها تركز على تطبيق الأدوات والأساليب على نطاق واسع في المؤسسات التعليمية على اختلاف مستوياتها. لكن كلا من مجالي تحليلات التعلم والتنقيب عن البيانات يعملان مع الأنماط والتنبؤ. وإذا أمكن التمييز في البيانات وكان لها معني ما يحدث، يمكن التنبؤ بما يجب أن يأتي لاحقا واتخاذ الاجراء المناسب.

وتظهر التطورات الحديثة في مجالي تحليلات التعلم والتنقيب عن البيانات حول الخصائص الديموغرافية في إطار أمثلة أعمال الطلاب (Reardon & Stuart, 2019). وعلي نفس المنوال، صارت الدراسات المنهجية وتعليم المعلمين تكتسب من الذكاء الاصطناعي لوحات العرض Dashboards التي تمكن المعلم تطبيق ردود فعل ديناميكية للممارسات أو القرارات التعليمية (Rosenberg et al, 2020). كما درس الباحثون في الولايات المتحدة شبكات صداقة الطلاب على نطاق أوسع وتحديد أثارها على سلوكيات الطلاب المحفوفة بالمخاطر الصحية (Harris, 2013) والعنصرية (Currarini et al, 2010). وقد

أدي ذلك لامتداد علم البيانات في الموضوعات التعليمية، الذي صار وليدا حاليا ويكتسب انتشارا واسعا بسرعة من ناحية أنواع البيانات المستخدمة، ويقدم بشكل متزايد فرصا وسبلا أكثر لفهم الهياكل والعمليات التعليمية.

ونفس أنواع البيانات التي تعلم نماذج المستخدم أو المتعلم يمكن استخدامها لتوصيف المستخدمين والتخطيط كما هو مستخدم، مما يعني تجميع المستخدمين المتشابهين في فئات باستخدام الخصائص البارزة، وإمكانية استخدام هذه الفئات لتوفير الخبرات لمجموعات المستخدمين، أو تقديم توصيات للمستخدمين، والتكيف مع كيفية أداء النظام.

لذلك، نمذجة المستخدم وتحديد سماته توحي بالتكيفات في الوقت الحقيقي. والتحليلات هي لأغراض تجريبية أكثر. وفي المقابل، بعض تطبيقات التحليلات والتنقيب عن البيانات هي لأغراض تجريبية أكثر.

كما أن نمذجة المجال تكون تجريبية لحد كبير وتهدف لفهم كيفية عرض الموضوع وإلى أي مستوي تفصيل. وأيضا تستخدم دراسة مكونات التعلم والمبادئ التعليمية التجربة لفهم ما هو فعال في تعزيز التعلم.

هذه الأمثلة تقترح أن الأفعال من التحليلات والتنقيب عن البيانات تكون آلية دائما، لكن هذا ليس هو الحال في كثير من الأحيان. حيث أن تحليلات البيانات المرئية تتضمن تواجد الإنسان عن كثب لفهم البيانات من اكتشاف النمط الأولي ونمذجة بناء لوحة معلومات متطورة التي تعرض البيانات بطريقة تمكن البشر التصرف بها.

وقد بدأت المدارس والمناطق التعليمية في تبني مثل هذه التحليلات على المستوي المؤسسي لاكتشاف مجالات التحسين التعليمي، وضع السياسات التعليمية، وقياس النتائج أيضا. وقد جعل ذلك أنشطة التعلم والتقييم مرئية تفتح الإمكانية للطلاب من أجل تطوير مهاراتهم في مراقبة تعلمهم، ورؤية كيف تعمل جهودهم علي تحسين فرص نجاحهم بشكل مباشر. ويكتسب المعلمون وجهات النظر الثاقبة عن أداء الطلاب، حيث تساعدهم في تكييف طرق تدريسهم فيما يرتبط ببدء التدريس، تناسب التكليفات وغير ذلك.

التطبيقات القوية لتحليلات التعلم وأيضا للتنقيب على البيانات تأتي مع كل من التكليفات المتطلبة والتحديات التعليمية التي يجب أن تفهم من قبل إدارات أو مراكز تكنولوجيا المعلومات التي تختص بجمع وتخزين البيانات المسجلة. بينما مطوري الخوارزمية سوف يتعرفون على التكليفات الحاسوبية التي تتطلبها أساليب تحليلات التعلم والتنقيب عن البيانات.

وأحد تحديات أنظمة البيانات التعليمية هو أنها غير قابلة للتشغيل البيني المتبادل، لذلك يتم جمع البيانات الإدارية معا، ويبقي مستوي بيانات الفصل الدراسي ممثلا تحديا للمطورين. ومع ذلك، الجمع بين هذه البيانات يمكن أن يعطي الخوارزمية المطورة قوة تنبؤية أحسن. كما تشكل كل من أداء الطلاب، التتبع علي الخط، الاختبارات المقننة، والاختبارات المنشأة من قبل المعلم صورة مبسطة لما قد يعرفه الطلاب التي قد تكون صعبة مما يحتم ان تلبي المعايير المقبولة للصلاحية، وتطلب عناية فائقة لخصوصية الطالب والمعلم والالتزامات الأخلاقية المرتبطة بالمعرفة والعمل علي بيانات الطالب.

٤. البيانات والطرق الجديدة

١/٤ البيانات الجديدة

قبل ظهور البحوث الكمية في التعليم المرتكزة على اعتبار البيانات الإدارية، على سبيل المثال، ملخصات الالتحاق بمقرر دراسي أو بيانات طولية من واجابات تكليفات الطلاب النهائية. وقد شهدت الأعوام الحديثة ارتفاعا ملحوظا في البيانات التعليمية الجديدة. وأحد الخصائص الرئيسة لهذه البيانات الجديدة أنها غير مهيكلة غالبا، أو إنها غير بديهية للهيكل في الشكل العلائقي. كما أن أحجام نصوص البيانات صارت كبيرة، تتضمن تداخلات وتسجيلات صوتية مسموعة ومرئية. علي سبيل المثال، يمكن دمج البيانات التعليمية الجديدة

في النماذج الحوسبية لإلقاء الضوء على الاتجاهات والحجم غير المسبوق بسرعة متناهية. إصدار هذه البيانات غير المهيكلة يعتبر أمرا بالغ الأهمية حيث أنها تشكل الغالبية العظمي لتوافر البيانات من أي منظمة بنسب تشغيل تصل لحوالي ٨٠٪ في الغالب الأعم (Shilakes & Tylman, 1998).

وقد حدد كلا من فيشر وآخرين (Fischer et al, 2020) أن كلا من التصنيف الجزئي وبيانات المستوي الكلي تمثلان إطار عمل لفهم البيانات غير المهيكلة. كما أن بيانات المستوي الكلي لها تعتبر مكونا زمنيا مرتبط بها أيضا وتتوافر من خلال البرمجيات المفتوحة المصدر للتعليم علي الخط MOOCs. أساليب المحاكات. المباريات والمعلمين الأذكياء حيث تنبثق بيانات التفاعل الدقيقة المرتبطة ارتباطا وثيقا بتفاعلات المعلمين. ويمكن التقاط البيانات الفردية من عينة معلمين كبيرة. والنقر فوق سجلات البيانات يمثل مثالا شائعا.

أما بيانات المستوي المتوسط لها مكون زمني محدود فإنها تضيف تعمق أكبر نحو تقييم قدرات المتعلمين المعرفية. الصفات الاجتماعية. العلاقات المستمرة. والظواهر المتعددة لبيانات النص كتلك التي يتم انتشارها من الوسائل الاجتماعية. منتديات برمجيات MOOC. أو النصوص الرقمية التي تم تغذيتها في طرق معالجة اللغة الطبيعية NLP التي تقع في هذا المجال بشكل كبير. كما تفعل تقارير الصداقة. الانتساب وسمات الشبكات الاجتماعية الأخرى.

أما مصادر بيانات المستوي الكلي التي تعمل على مستوي المؤسسة التعليمية الكلي يمكن مشاهدتها كما في حالة مستودعات المياه Watershed بين مجموعات البيانات المهيكلة في أواخر تسعينيات القرن العشرين الماضي. والبيانات غير المهيكلة المنشأة من منصات رقمية ديناميكية الطابع التي لا يمكن تضمينها معلومات ثابتة فقط. كالتركيبة السكانية للطلاب. لكن أيضا كمكونات ديناميكية مثل الحضور الأسبوعي. الارتباط. ودرجات الإنجاز (Fischer et al, 2020). كما يمكن الاستفسار عن نمو مصادر البيانات الجديدة كنمو لعلم بيانات التعليم. أما البيانات الصوتية المسموعة والمرئية (التي تكون أبعد من مجرد نصوص مكتوبة) يمكن تحليلها أيضا للنغمة الضمنية ولغة الجسد الممكن مقارنتها عندئذ مع النتائج المستمدة من النص المكتوب فقط.

وهناك جهود ممكن مشاهدتها في علم بيانات التعليم لربط مجموعات بيانات متعددة معا للإبلاغ عن أسئلة البحوث عبر المجالات الموضوعية. أو لتحديد تشابهات المجال عبر منصات التعلم الرقمي (Li et al, 2021). وفي هذا السياق ربط ريردون Reardon بيانات الإنجاز التعليمية مع بيانات ضرائب الدخل لتسليط الضوء حول اتجاهات عدم مساواة الدخول للمواطنين في الولايات المتحدة (Readon, 2016). كما أن كلا من فيجليو ولوكاس (Figlio & Lucas, 2004) ربطا البيانات من بطاقات التقارير المدرسية مع أسواق الإسكان لفهم ما إن كانت درجات بيانات التقارير المدرسية تؤثر علي المواقع السكنية للأسر وأسعار المنازل.

وقد بدء البحث في التعليم العالي يستخدم أيضا الذكاء الاصطناعي لربط نقل مسارات المقررات التعليمية عبر المعاهد (Pardas et al, 2019). وربط معلومات واسعة النطاق عن أعضاء هيئة التدريس لمنح براءات الاختراع (Manijunath et al, 2021). ومساندة تدريب الخريجين عن طريق منحهم منحا لتأكيد عوائد سوق العمالة في الاقتصاد الأوسع (Weinberg et al, 2014). والان توجد إمكانيات للباحثين لدمج البيانات التي تم جمعها من المسوح مباشرة مع بيانات أخري مفتوحة المصدر أو غير ذلك.

٢/٤ الطرق الجديدة

نمو علم بيانات التعليم السريع يرتبط بالنمو في تطبيقات التعلم الآلي Machine Learning. ومن بين خوارزميات التعلم الآلي. خوارزميات التعلم المشرف عليه Supervised Learning كتطبيق شائع في مجال

التعليم (Olive et al,2020). وتشتمل خوارزميات التعليم المشرف عليه على متغيرات مستقلة يتم التنبؤ بها من مجموعة متغيرات مستقلة. وينشئ التعلم الآلي وظيفة تحدد المدخلات للمخرجات المرغوبة باستخدام كلا من مجموعات المتغيرات. وأمثلة من تلك الخوارزميات التي تتضمن الانحدار، تطبيقات الشبكات العصبية العميقة DNN، خوارزميات أقرب الجيران K-Nearest Neighbors، أشجار القرار Decision Trees، والغابة الجزافية Random Forest. أما خوارزميات التعلم غير المشرف عليه Unsupervised Learning فقد تم تطبيقها أيضا (Zhang et al, 2017). وحيث أننا لا نملك متغيرات تابعة فغالبا يتضمن المدخل التجمع Clustering أي عدد من الأفراد معين في مجموعات قائمة على التشابه. كما تم تطبيق التعلم المعزز Reinforcement Learning أيضا للتعليم في أعمال متعددة (Bassen et al, 2020). ويتضمن التعلم المعزز المدخل الذي يتعرض لبيئة، حيث تدرب الخوارزمية نفسها من خلال تكرار التجربة والخطأ المتعدد الأقرب لما نشاهده في عملية قرار ماركوف Markov. وعلي ذلك يمكن مشاهدة عاملين يساهمان للنمو الهائل للتعلم الآلي في التعليم، هما:

أولا: الانفجار الهائل في البيانات النصية والمتسلسلة التي تستلزم مدخلات منهجية مناسبة لاستيعاب تحليلها (Pardos, 2017)، وذلك من خلال ارتفاع شعبية برمجيات المقررات مفتوحة المصدر علي الخط MOOCs والتعلم الإلكتروني لأحجام بيانات ضخمة تم إنتاجها من خلال أنشطة التدريس والتعلم في المقررات الدراسية المتاحة علي الخط (Kizileec & Brooks, 2017). حيث أن أحجام البيانات الضخمة هذه تعتبر ضرورية لتحقيق مكاسب في الأداء من خلال خوارزميات التعلم الآلي. وتنشئ مجموعات البيانات الضخمة أيضا أسرة الاختبار للتداخلات التي تحتاج بيانات تدريب جوهرية وتصميم دقة تنبؤ خوارزميات التعلم الآلي.

ثانيا: طمس الحدود الانضباطية يمكن علماء علم الحاسب الآلي والاحصائيين الموهوبين لتطبيق مهارات علم البيانات نحو مخاطبة التحديات المجتمعية الملحة. وقد رعت الجامعات هذه الاتجاهات بشكل كبير مع مناصب أعضاء هيئة التدريس وإنشاء برامج الزمالة كما في حالة علم الحاسب الآلي بالإضافة للصالح الاجتماعي وعلم البيانات من أجل زمالة الصالح الاجتماعي.

وفيما يتصل بمجال «معالجة اللغة الطبيعية NLP» كمجال فرعي للتعلم الآلي الذي يتطلب اهتماما خاصا، فقد تم تطبيقه لأقصي حد لبيانات النص في المواقف التعليمية التي إما أنها جمعت مباشرة أو تم نسخها من تسجيلات الوسائط المتاحة فعليا. ويمكن تطبيق أساليب معالجة اللغة الطبيعية على نطاق واسع للنصوص الكبيرة في التعليم لفهم سمات مثل المشاعر الضمنية في النص، والخبرة في المعلومات المقدمة، والموضوعات المحددة بواسطة نمذجة الموضوع (Lucy et al, 2020). ويمكن تقييم خوارزميات معالجة اللغة الطبيعية بطريقة ديناميكية في الفصول الدراسية فيما يتعلق بإجادة القراءة للطالب وإنشاء ردود فعل في الوقت الفعلي من أجل التحسين (Li et al, 2021). وقد صارت خوارزميات معالجة اللغة الطبيعية تستخدم لتقديم ردود فعل قابلة للتنفيذ حول النثر، النحو وآليات الكتابة العامة مع السماح بفحوصات لها (Shum et al, 2016). علي سبيل المثال، التوقيع المحتمل للفئة المعينة في عناصر المواد التعليمية المكتوبة (Alvero et al, 2021). بالإضافة إلى مكونات مواجهة الطالب، منصات معالجة اللغة الطبيعية فإنها تقدم مكونات تواجه المعلم أيضا، مما يساعد المعلمين إجراء تقييمات تكوينية قوية، وقد يكون الأمر بخلاف ذلك صعبا في الفصول الدراسية مع نسب كبيرة للطلاب والمعلمين (Burstein et al, 2014).

فئة أخري جديدة من الطرق التي تتم تمكينها بواسطة مراكز علم بيانات التعليم حول تحليل الشبكات الاجتماعية تتمثل في أن وسيلة التعليم هي التواصل مما يشكل العلاقات الاجتماعية التي تتأثر بدورها بواسطة العلاقات المنشأة التي تقود السلوكيات الشخصية واتجاهات المشاركين في المدرسة. وتعتبر

الطرق التحليلية للشبكة كمجال عرض هذه العلاقات. التداخلات والتأثيرات الشخصية الناشئة في هذه المواقف التعليمية والمنصات المتاحة على الخط. وركز تحليل الشبكة منذ فترة طويلة على العلاقات المباشرة بين منتفعي العملية التعليمية مثل الطلاب في المدارس (McPherson et al, 2001)، مدرسي الفصول الدراسية وغيرهم. إلا أنه تم توسيع نطاقها عبر طرق حوسبة جديدة التي تقلل من تعقد الارتباط بين الأشخاص إلى فضاءات ذات أبعاد عدة (Grover & Leskovec, 2016). بالإضافة لذلك، مجموعة الطرق الإحصائية الاستنتاجية تطورت للتنبؤ إلى جانب تطوير شبكات النماذج كعلاقات مباشرة وهياكل تابعة. وكثير من هذه الطرق الجديدة طورت مع بيانات المدرسة والفصل الدراسي كحالات اختبار يمكن ايجادها مرتبة على نطاق واسع في قضايا «مجلة الشبكات الاجتماعية Journal of Social (Networks» (Cranmer et al, 2020). وتنادي طرق الشبكات الاجتماعية بوجهات النظر البيئية للتعلم، كما تشاهد بشكل جيد لقواعد البيانات العلاقية وإلى المعلومات عن الانتماءات المختلفة، والتفاعلات الممثلة بشكل شائع في السجلات التنظيمية ومنصات الويب. باختصار، البيانات الجديدة والطرق الجديدة نشأت كقوي تعزز بعضها البعض تؤجج نمو علم بيانات التعليم السريع وكمصدر أكثر للمعلومات غير المهيكلة التي تصبح في متناول البشر. ومن المتوقع رؤية طرق علم بيانات التعليم تتطور بسرعة لمطابقة هذه التحديات.

فمثلا الطرق المعرفية للاقتصاديين وجدت في بعض الأحيان توترات مع مجتمع التعلم الآلي في مجال التنقيب عن البيانات، لكن في التعليم العالي بدأت نظم علم بيانات التعليم تقديم بعض المنظورات التكميلية علي البيانات المؤسسية والطلابية (Chaturapruek et al, 2021)

٥. مفهوم علم البيانات وتطبيقاته في التعليم

١/٥ مفهوم علم البيانات التعليم

من خلال ما تقدم يمكن تصور علم بيانات التعليم كالتقاط مجموعة أوسع من البيانات والطرق من خلال مراجعة الآداب المنشورة عنه، بينما التنقيب عن بيانات التعليم وتحليلات التعلم ترتبطان ارتباطا وثيقا مع الدراسات الإنمائية والسيكولوجية في التعليم. أما الاتجاهات الحديثة في علم بيانات التعليم فقد شهدت تطبيقا أوسع. وبدأ علم بيانات التعليم يؤدي دورا في دراسة موضوعات ذات ترتيب أعلي كما في السياسات والتنظيمات، بينما يشاهد أيضا تطبيقا مستمرا في دراسة درجات اختبار عالية المخاطر ومقاييس قراءة دقيقة.

كما يمكن تصور علم بيانات التعليم، بمعناه الضيق، كتطبيق الأدوات ووجهات النظر من علم الإحصاء وعلم الحاسب الآلي للظواهر والمشكلات التعليمية. ويمكن مناقشة ذلك أيضا لمزيد من التوسع، حيث أن علم بيانات التعليم يمثل مظلة لمجموعة جديدة وطرق كمية غير تقليدية (مثل التعلم الآلي، تحليل الشبكات ومعالجة اللغة الطبيعية) التي تطبق للمشكلات التعليمية المستخدمة بيانات جديدة غالبا.

ويسهم ذلك في اكتشاف وتأكيد أن البيانات الجديدة و/أو الطرق المستخدمة بواسطة مناقشة أنواع بحوث علم بيانات التعليم تمكن من خلال كافة الإمكانيات الناشئة. والعرض التالي يلقي الضوء على أبعاد كلا من البيانات الجديدة والطرق الجديدة المستخدمة في علم بيانات التعليم.

٢/٥ تطبيقات علم بيانات التعليم في التعليم

١) المهارات العاطفية الاجتماعية:

المهارة العاطفية الاجتماعية تمثل مجالا مهما وتحتاج للتطوير من خلال العملية التعليمية. ومن خلالها يتعلم الطفل اكتساب القدرة على الفهم، التحليل، التعبير، وإدارة مشاعره. كما يتعلم الطفل أيضا كيفية تطوير علاقاته مع الآخرين. وتسهيل النمو في المهارة العاطفية الاجتماعية يعتبر مهمة هامة للمعاهد التعليمية علي اختلاف مستوياتها وأنواعها ويمثل ذلك مثلا لمهارة غير

أكاديمية التي تلعب دورا رئيسيا في تفسير قدرات تعلم الطلاب.

وقد وجد سابقا مسوحا إحصائية كثيرة لتقييم هذه المهارة العاطفية الاجتماعية. ومع التطورات في منهاجيات الحوسبة، صار ممكنا جمع كميات بيانات ضخمة. كما أنه مع نماذج اكتشاف المعرفة المشكلة في علم البيانات وأساليب التنقيب عن البيانات وتحليلات التعلم صار ممكنا لجمع مثل هذه المعلومات الضخمة وتضمينها في الأدوات المتواجدة لإنتاج نتائج أحسن. إلي جانب ذلك، مع البيانات المستخلصة، تمكن علماء البيانات تطبيق أساليب تحليلية تنبؤية عديدة لمساعدة المدرسين في فهم دافعية الطلاب لدراسة المقرر التعليمي المعين.

٢) مراقبة متطلبات الطلاب:

توجد أساليب تعليم وتقييم عديدة الموظفة بواسطة المؤسسات التعليمية، لكنها مثل المنهجيات التقليدية لم تتمكن حتى لالتقاط وتغليف كل اتجاهات ومنصات خدمات الطلاب المهمة. إضافة لذلك، معظم أساليب التقييم لم تكن جاهزة في الوقت الحقيقي. ومع التطورات الحديثة في تحليلات البيانات الضخمة، صار ممكنا للمدرسين لفحص الطلاب الخاصة بهم بناء على أدائهم والمراجعات المتعلقة بهم. وكنتيجة مراقبة متطلبات الطلاب، صار المعلمون غير قادرين على تقديم الردود المناسبة وحتى تغيير منهاجيات تدريسهم تلبية لتوقعات الطلاب. ووجد في مرات عديدة تحيزا غير واع لدي المعلمين تجاه بعض الطلاب. أما من خلال منصة البيانات فإنها سوف تعامل مستخدميها مع عدم التحيز. كما أنه لن يكون هناك أي تحيز في تقييم أداء الطلاب. وهذا سوف يقدم منصة متساوية لكل الطلاب للمشاركة وتطوير مهاراتهم.

٣) ابتكار المنهج الدراسي:

يجب ا، تواكب كثير من الجامعات متطلبات الصناعات المختلفة وذلك لتقديم المقررات الدراسية المناسبة لطلابهم. علاوة على ذلك، فإن ذلك يمثل تحديا للجامعات لمواكبة نمو الصناعات. ومن أجل استيعاب هذا، صار لزاما على الجامعات استخدام نظم علم البيانات لتحليل الاتجاهات المتنامية في السوق. ومع استخدام المقاييس الإحصائية العديدة وأساليب المراقبة، يمكن لعلم البيانات أن يكون مفيدا لتحليل الأنماط الصناعية ومساعدة مطوري المقررات الدراسية لاستيعاب موضوعات مفيدة. إضافة لذلك، استخدام التحليلات التنبؤية صار يمكن الجامعات من تحليل الطلبات لمجموعات المهارة الجديدة، وإنشاء المقررات الدراسية لمخاطبتها.

4) قياس أداء المعلمين:

٣/٥ دراسة حالة علم البيانات التعليم:

يعتبر التعليم العالي بوابة لحياة ناجحة لكثير من الطلاب في التعليم العالي علي أمل الحصول على وظيفة. علي أي حال، كثير من الطلاب يتسربون من التعليم بسبب التحديات الأكاديمية أو المالية. وطبقا لدراسة مسحية حديثة في الولايات المتحدة الأمريكية لجامعة فلوريدا وجد أن ٥٤،٨ ٪ من الطلاب الملتحقين سيكونون غير قادرين للانتهاء من دراجتهم الأكاديمية في مدي 6 سنوات. هذه النسبة المذهلة تعني أن الصناعات سوف يكون لها وصولا محدودا من الأفراد المؤهلين مما قد يعوق النمو الوطني أيضا. ومن أجل حل هذه القضية التي تهدد المؤسسات التعليمية تم التطلع نحو علم بيانات التعليم. وقد وظفت جامعة فلوريدا برنامج IBM InfoSpher لاستخلاص النتائج ونقل البيانات من موارد متعددة مع استخدام برنامج IBM SPSS Moduler للتحليلات التنبؤية ونمذجة البيانات (Cheng et al, 2015). وارتبط ذلك مع منصتين ومع برنامج IBM Cognos Analytics الذي هو أداة قوية لذكاء الأعمال مبني على الويب ويقدم مجموعة أدوات عديدة للإعلام، التحليل، ومراقبة الأحداث خلال التصور التفاعلي. واستخدام هذا البرنامج يمكن الجامعات لتحليل والتنبؤ بأداء الطلاب. حيث يستخدم متغيرات عديدة كخلفية الطالب، ووضعه الاجتماعي وخلافه من العناصر الأخرى التي

تسهم في تقييم إمكانية التسرب المحتملة للطالب. وعلي ذلك، يمكن مساعدة الجامعة في رسم سياساتها وتقديم تداخل مبكر للطلاب الذين على وشك التسرب.

وقد أنشأت شركة IBM مشروعا للمداري لتقييم الطلاب بناء على أدائهم من خلال التحليلات. علي سبيل المثال، استفادت جامعة فلوريدا من تحليلات الشركة من خلال برنامج IBM Cognos Analytics لتتبع أداء الطلاب وعمل التنبؤات الضرورية ، كما أن الجامعة تستخدم برمجيات MOOCs ومنصات التعليم علي الخد لغلم البيانات من أجل تتبع أداء الطالب ، ولآلية تقييم الواجبات والتكليفات الدراسية وتحسين المنهج الدراسي المبني علي ردود فعل الطلاب.

٦. فرص وتحديات علم بيانات التعليم

هناك كثير من الفرص والتحديات المتعلقة بعلم بيانات التعليم والتي منها على سبيل المثال أنه يجب علي علم بيانات التعليم الاعتماد على مجموعة التقاليد الغنية التي تؤدي إلي إعلام البحوث التربوية وخاصة ما يتعلق بتقاليد العلوم الإنسانية والاجتماعية (McFarland et al, 2015). ومجالات التقاليد هذه منخرطة في مستجمعات المياه المماثلة Watershed، حيث أن البيانات الضخمة Big data وعلم البيانات يتداخلان معا ويحدثان ثورة في مجالهما. والتقائهما يكشف التوترات والنجاحات التي ربما لعلم بيانات التعليم التعلم منهما. حيث أن هناك مدخل واحد محتمل لتكامل علم البيانات والتعليم على نفس المنوال. ويجب على المشروعات في علم بيانات التعليم أن تهدف لدمج أفضل التقاليد المنهجية المتأصلة في العلوم الأخرى جنبا إلي جنب مع تقاليدهم النظرية والمفاهيمية. حيث تصبح هه البيانات متاحة بسهولة أكبر، كما يمكن استخدام التعلم الآلي لتجميع وفهم السياقات الفنية لاجتياز الطب في جميع مساراتهم بعد الثانوية العامة (Pardos et al, 2019)، وفتح أفاق جديدة للتدخل في استخدام النماذج التنبؤية (Bird et al, 2021). مقدمة مداخل علم البيانات لا تحتاج مجالات التخلص من السلف التي تحققت عن جدارة. كما أن المبادئ الراسخة التي ليست حالة الاستعمار بين المجالات. مثال علي ذلك، التصاميم التجريبية من تخصص الاقتصاد يمكن أن تطبق كمعيار ذهبي لتقييم آثار تداخلات علم بيانات التعليم المستنيرة في التعليم العالي في بعض الأحيان.

مثال آخر، يتصل بما يعادل قرن من الزمان من قيمة البحوث السيكومترية تقدم مداخل نمذجة عديدة قد يكون من المفيد الاستفادة منه جنبا إلي جنب مع مداخل التعلم الآلي الحديثة. المداخل السيكومترية قد تكون مفيدة علي حد سواء من أجل التقدم المنهجي اللاحق (من حيث الميزات التي تستحق الاهتمام)، وتكون مفيدة أيضا كمقاييس للطرق الجديدة من علم البيانات. مثل هذا المنظور يكشف أن المكاسب من مداخل التعلم الآلي غالبا ما تكون ثانوية نسبيا. ويحتم على علماء البيانات للاستماع والتعلم من العمل الدقيق على أخذ العينات وانظر فيها عندما تون جميعها غالبة في قرن كامل، وتعكس البيانات المجودة لعدد الجمهور المعروف (McFarland & McFarland, 2015). وكثير من الصدمة والرعب من علم البيانات يجب تطبيعها وجعلها أكثر فائدة عندما نتذكر الرؤي والبصائر الي تنقلها مجالات أخري..

وأحد الأسئلة المهمة يتمثل في كيف تدرب الطلاب أحسن لدخول ميدان علم بيانات التعليم. ومن وجهة نظرنا، يجب أن يكون هناك تركيزا واضحا على المشكلات ذات الصلة للتعليم التي يمكن تتبعها في ضوء جمع البيانات التي هي في متناول أيدينا؟ وعلي الرغم من حدوث انفجار مبني للبيانات في التعليم، لا يزال لدينا بيانات أقل بكثير (بينما تتواجد بيانات غنية بالعديد من المجالات) أي مما هو متاح في المجالات الأخرى. وقد حد ذلك من قابلية التطبيق لمعظم مداخل الخوارزمية (Bird et al, 2021). كما يجب أن نعمل أيضا علي بناء اخلاقيات التناسب بين الطلاب.

بينما الكثير في التكنولوجيا اتخذت خطوة نحو التحرك بسرعة وكسر صدي الأشياء. ويمكن التفكير في هذا الاتجاه بأنه غير ملائم بشكل كبير نظرا لطبيعة التعليم (المرتبط بتنوع أصحاب المصلحة والعناية المحتاج إليها عند التعامل مع القضايا المؤثرة على الطلاب الشباب). وبدلا من ذلك، هناك حاجة لأن نكون شبيهين بالأطباء مع قسم ابقراط Hippocratic. وقضايا العدالة علي سبيل المثال، لا يمكن اعتبارها في النهاية بل بدلا من ذلك تحتاج أن تكون مركزية من الخارج.

وبذلك يمكن عرض بعدين يتصلان بالحذر الكامل:

أولا، يتعلق هذا البعد بحدود علم بيانات التعليم المرتبط بالمدخل الحاسوبية المركزة في علم بيانات التعليم، وتعتبر مثيرة فيما تقدم من رؤي وبصائر جديدة عما كانت عليه في المشكلات القديمة، أو السماح لأنواع جديدة من البيانات ووجهات النظر لدخول البحث التربوي. لكن هذه البيانات والمداخل لن تكون حلا سحريا. ويمكن رؤية الأساليب الحاسوبية في النواقص الحديثة (من فئات التحدي الهشة (Salganik et al, 2016).

وفي هذا الإطار، إضافة البيانات الثرية وأساليب النمذجة المتطورة لم تزداد القدرة على تنبؤ العديد من نتائج دورة الحياة ذات الصلة في دراسة الشباب بشكل كبير. هذه النتائج مفيدة من حيث تحديد التوقعات في كل من علم السلوك والعلم التربوي حيث أنهما يمثلان تحديا بشكل عام. معظم الابتكارات عن البيانات أو جانب الحوسبة الذي ينبغي توقعه لجلب تحسينات هامشية في فهمنا فقط.

ثانيا، البعد الثاني يتعلق بالمشكلات الخبيثة الخاصة بالتحيز في مداخل الحوسبة، والحاجة للعمل نحو الانصاف. علي الرغم من المراجع المبكرة لاحتمالات التحيز في الحوسبة (Friedman & Nissenboum, 1996)، التي تمثل أثارا ضارة درست بدقة في السنوات القليلة الماضية واعتبرت ظاهرة مترابطة للغاية مع ظهور الآلية والذكاء الآلي. كما وجد أن هناك أربع مشكلات محتملة حول التحيز في مداخل علم بيانات التعليم التي يحتم على الباحثين الانتباه لها جيدا (Zou & Schiebinger, 2021) وهي:

أولا: أغلبية الخوارزميات الرقمية النامية مبنية على البيانات المولدة من مجموعة تدريب من المستخدمين السابقين، وغالبا يكونوا غير ملائمين لصالح مجموع المستخدمين. وقد ينتهي هذا بمزيد من تهميش المعرفة المولدة بواسطة الفئات الاجتماعية المحرومة تقليديا.

ثانيا، توجد مشكلة مع طبيعة التنبؤ ذاته الضمنية في معظم خوارزميات علم البيانات. حيث تنظر التنبؤات إلي الماضي لعمل التخمينات نحو أحداث المستقبل. كما يمكن أن تعرض طرق التنبؤ عدم المساواة في الماضي في المستقبل أيضا.

ثالثا، كثير من خوارزميات الوقت الحالي تعمل بسرعة وحجم غير مسبوقين. فمثلا برنامج Google Translate يخدم أكثر من ٢٠٠ مليون مستخدم يوميا (Prates et al, 2020). حيث أن الوظائف التي كانت مكلفة سابقا وبطيئة وتعتبر واحد لواحد يمكن حاليا جعلها آلية لتصبح أرخص وأسرع وتخدم جماهير أكبر حاليا. وهذا يمكن بالتأكيد كثير من الناس في الاستفادة من الخوارزميات الآلية. لكن ترجمة نظام متحيز ي مكن أن تخدم ما يزيد علي ٢٠٠ مليون كلب بحث متحيز (Prates et al, 2020).

رابعا، نقص السيطرة البشرية على ما يذهب داخل البرمجيات الآلية يعتبر أحيانا مؤشرا على الحياد. هذا الافتراض قد يمثل إشكالية لأن العدالة تعتبر غير متأصلة في أي خوارزمية، وهي بالأحرى تمثل جودة يجب تصميمها بعناية فائقة من أجل صيانتها. وعدم الاهتمام بالعدالة يمثل مخاوف في علم بيانات التعليم الذي يمكن أن يسبب ضررا (عندما تعزز الخوارزمية تنقية المعلومات على امتداد خطوط الكيان)، والتخصيص (عندما تخصص الخوارزميات، أو حجب فرصة مجموعات معينة أو مورد معين (Crawford, 2017). كما حذر (Susskind,

2018) في مجلة سياسات المستقبل Future Politics التي تنشر بواسطة جمعية البحث التربوي American Educational Research Association (AERA)وتتحكم في تدفق المعلومات في مجتمع يمكن الفرد فيه التأثير على احساسه م المشترك بالحق أو الزيف، عادل أو غير عادل، نظيف وغير نظيف، وعلي ما يبدو لائق وغير لائق، حقيقي وغير حقيقي، صحيح وخطأ، معروف وغير معروف

باختصار، علم بيانات التعليم يقدم فرصا لإبداع وتقدم البحث التربوي. وعلي ذلك، فإن مجلة AERA Open تتضمن موضوعات خاصة عن علم بيانات التعليم تكشف وجهات النظر والاهتمامات التي عقدتها مجموعة فرعية من العلماء المهتمين بالمشاركة المباشرة مع البحوث التربوية، كما تقدم أيضا لمحة عامة من الجناح الأوسع للبيانات والطرق التي تشكل علم بيانات التعليم. وبينما يتواجد لدينا درجة من التفاؤل عن أسئلة البحث سيكون قابلا للتتبع نظر للإمكانيات المتاحة لعلم بيانات التعليم، كما يمكن الاعتراف بالحاجة إلي جرعة صحية من الواقعية والحذر أيضا. حيث يدخل العلماء الجدد في هذا المجال البحثي الناشئ ويصبحوا مشاركين علم بيانات التعليم وعليهم عمل ذلك وأعينهم مفتوحة حتى يتمكنوا من تحقيق أفضل استخدام وتطبيق هذه المداخل الإبداعية.

٧. الاستنتاج والتوصيات:

علم بيانات التعليم وما يتضمنه من تنقيب عن البيانات وتحليلات التعلم صار لديه القدرة على جعل البيانات التي كانت مختفية وغير ملحوظة من قبل مرئية وقابلة للتنفيذ، من أجل المساعدة في تعزيز المجالات واكتساب القيمة من تطبيقاتهم العملية.

ويتضح مما سبق عرضه يمكن تفهم الدور المتنوع لعلم البيانات في مجال التعليم، حيث قد أدي لثورية كل قطاع وإحداث تغييرا إيجابيا للسيناريوهات التعليمية. وعلي ذلك، اصبح علم بيانات التعليم أداة فعالة لتشكيل العليم لا فقط المهارات الأكاديمية، لكن أيضا المهارات غير الأكاديمية كالمهارات العاطفية الاجتماعية. إضافة لذلك، يستخدم أيضا لتقييم أداء الطلاب والمعلمين. وبذلك صارت الجامعات تستخدم علم بيانات التعليم لتنسيق المنهج الدراسي بها، والأخذ في الاعتبار اتجاهات النمو في الصناعات من أجل النمو الوطني المنشود.

وبذلك يمكن التوصية للمعلمين والإداريين بما يلي:

- تطوير ثقافة استخدام البيانات لاتخاذ قرارات تعليمية.
- تضمين إدارات أو مركز تكنولوجيا المعلومات في التخطيط لجمع البيانات واستخدامها.
- أن يكونوا مستهلكين أذكياء يطرحون أسئلة حرجة عن العروض التجارية وإنشاء الطلب للأوجه والاستخدام الأكثر فائدة.
- البدء مع المجالات المركزية التي سوف تساعد البيانات، تظهر النجاحات، والتوسع في مجالات جديدة.
- التواصل مع الطلاب وأولياء الأمور عن من أين تأتي البيانات وكيف تستخدم.
- المساعدة في ملاءمة سياسات الدولة مع المتطلبات الفنية لنظم العليم على الخط.

كما يمكن توصية الباحثين ومطور البرمجيات بالتالي:

- إجراء البحوث حول سهولة واستخدام عروض البيانات.
- مساعدة المدربين والمعلمين لكي يكونوا أكثر فعالية في حجرات الدراسة مع المزيد من القت الفعلي والأدوات المساندة للقرارات المبنية على البيانات.
- مواصلة البحث والتقصي عن الطرق لاستخدام وتحديد معلومات الطلاب لأنها سوف تساعد أكثر في كشف البيانات المجهولة عنهم عند الحاجة، وفهم كيفية توافق البيانات عبر الأنظمة التعليمية المتنوعة والمختلفة.

- تفهم كيفية إعادة تحديد أعراض النماذج التنبؤية المطورة في سياق واحد إلي آخر.
- إنشاء تعاون قوي راسخ ومستمر عبر قطاعات البحوث التربوية والتجارية للشركات التجارية التي تعقد دورات تطوير سريعة حيث يمكنها توليد بيانات مفيدة للبحوث للمساعدة في نطاق المناطق التعليمية والمدارس ومؤسسات التعليم العالي التي تريد بيئة تعليمية فحصت بشكل صحيح.

المراجع

1. Alvero, A. J. et al (2021). Essay Content is Strongly Related to Household Income and SAT Scores: Evidence from 60000 Undergraduate Applications; CEPA Working Paper. https://cepa.stanford.edu/SH_es/default/files/wp21-c3-vo4202.pdf

2. Baker, R. S. (2013). "Educational Data Mining: Potentials and Possibilities," Presented at: American Educational Research Association Annual Meeting, Francisco: CA.

3. Bassen, J. et al (2020). "Reinforcement Learning for Adaptive Scheduling of Educational Activities," 2020 CHI Conference on Human Factors in Computing Systems Honolulu, HI

4. Bird, K. A. et al (2021). "Bringing Transparency to Predictive Analytics: A Systemati Comparison of Predictive Modeling Methods in Higher Education," AERA Open, Vol. 7, No. 1, https://doi.org/10.1177/23328584211037630.

5. Brackett, M. A. et al (2015). Enhancing Academic Performance and Social and Emotional Competences wit the RULER Feeling Words Curriculum," Learning and Individual Differences, Elsevier, pp. 1-7.

6. Burstein, J. et al (2014). "From Teacher Professional Development to the Classroom: How NLP Technology can Enhance Teachers' Linguistic Awareness to Support Curriculum Development for English Language learner," Journal of Educational Computing Research, Vol. 51, No. 1, pp. 119-144. https://doi.org/10.2190/EC.51.1.f

7. Chaturapruek, S. et al (2021). "Studying Undergraduate Course Consideration at Scale," AERA Open, Vol. 7, No. 1, https://doi.org/10.1177/2332858421991148.

8. Cheng, Q. et al (2015). "Mining Research Trends with Anomaly Detecting Models: The case of Social Computing Research," Scietometrics, Vol. 103, No. 2, pp. 453-469.

9. Cranmer, S. J. et al (2020). Inferential Network Analysis, Cambridge University Press

10. Crawford, K. (2017). "The Trouble with Bias," In: Conference on Neural Information Processing Systems, Long Beach, CA

11. Currarini, S. et al (2010). "Identifying the Roles of Race-Based Choice and Chance in High School Friendship," In: Networks Formation Proceedings of the Nation Poraayk-Academy of Sciences, Vol. 107, No 11, pp. 4857-4861.

12. Donoho, D. (2017). "50 Years of Data Science," Journal of Computational and Graphical Statistics, Vo. 26, No. 4, pp. 745-766.

354-363.

20. Lucy,L. et al (2020). "Content Analysis off Textbooks via Natural Language Processing: Findings on Gender, Race and Ethnicity in Texas US History Textbooks," AERA Open, Vol. 6, No.3. https://doi.org/10.1177/332858420940312

21. Manjunath, A. et al (2021). Comprehensive Analysis of 2.4 Million Patent-to-Research Citations Maps the Biomedical Innovation and Translation Landscape. Nature Biotechnology, Vol. 39, No. 6, pp. 678-684. https://doi.org/10.1038/41587-021-00940-5.

22. McFarland, D. A. et al (2015). "Sociology in the Era of Big Data: Ascent of Forensic Social Science," The American Sociologist, Vol. 47, No. 1, pp. 12-35. https://doi.org/10.1007/s12108-015-9291-8

23. McFarland, D. A. & McFarland, H. R. (2021). "Big Data and Danger of Being Precisely Inaccurate," Big Data and Society, Vol. 2, No. 2, https://doi.org/101177/20539551715602495.

24. McPherson, M. et al (2001). Birds of a Feather: Homophily in Social Networks," Annual Review of Sociology, Vol. 27, pp. 415-444.

25. Olive, D. M. et al (2020) "A Supervised Learning Framework: Using Assessment to Identify Students at Risk of Dropping out of a MOOC," Journal of Computing in Higher Education, Vo. 32, No. 1, pp. 9-26. https://doi.org/10.1007/s12528-019-09230-1

26. Pardos, Z. A. (2017). "Big Data in Education and the Models that Love them," Current Opinion in Behavioral Sciences, Vol. 18, pp. 107-113.

13. Figlio, D. N. & Lucas, M. E. (2004). What's in a Grade? School Report Cards and the Housing Market," American Economic Review, Vol. 94, No. 3, pp. 591-604. https://doi.org/10.1257/000282041464489.

14. Fischer, C. et al (2020). Mining Big Data in Education: Affordances and Challenges," Review of Research in Education, Vol. 44, No. 1, pp.130-140. https://doi.org/10.3102/0091732x20903304

15. Ftiedman¸B. & Nissenboum, H. (1996). "Bias in Computer Systems," ACM Transaction Information Systems, Vol. 14, No. 3, pp. 330-347. https://doi.org/10.1145/230538.230561

16. Grover, A. & Leskovec, J. (2016). "nod2vec: Scalable Feature Learning for Networks," In: 22nd ACM SIGDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA.

17. Harris, K. M. (2013). The Add Health Study: Design and Accomplishments. University of North Carolina, Chapel Hill. https://doi.org/10.17615/c6tw87

18. Kizileec, R. F. & Brooks, C. (2017). "Diverse Big Data and Randomized Field Experiments in MOOCs,"In: Lang, CC. et al (eds.) Handbook of Learning Analytics, Society for Learning Analytics Research, pp. 211-222.

19. Ii, Z. et al (2021). "Learning Skill Transfer Models Across Systems," In: Dowell, N. et al (eds.) Proceedings of the 11th International Conference on Learning Analytics and Knowledge, ACM, pp.

Mass Collaboration, "In: Proceedings of the National Academy of Science of U.S.A,., Vol. 117, No. 15, pp. 8398-8403. https://doi.org/10.1073/pnas.19915006117.

34. Shilakes, C. C. & Tylman, J. (1998). Enterprise Information Portals. Merril Lynch

35. Simens, G. (2013). "Learning Analytics: The Emergence of a Discipline," American Behavioral Scientist, Vol. 57, No. 10, pp. 1380-1400. https://doi.org/10.1177/0002764213498851.

36. Shum, S.B. et al (2016). "Critical Perspectives on Writing Analytics," In: Six International Conference on Learning Analytics and Knowledge. Edinburgh, Scotland, UK (Conference Session)

37. Susskind, J. (Future Politics: Living Together in a World Transformed by Tech. , Oxford University Press

38. Weinberg, B. A. et al (2014). "Science Funding and Short-Term Economic Activity," Science, Vol. 344, No. 6179, pp. 41-43. https://doi.org/10.1126/sscience.1250055.

39. Zhang, N. et al (2017). "Characterizing Students' Learning Behaviors Using Unsupervised Learning Methods," In: Andere, A. et al (eds.) Proceedings of the International Conference on Artificial Intelligence in Education, pp. 340-441.

40. Zou, J. & Schiebinger, L. (2021). "Ensuring that Biomedical AI Benefits Diverse Population," EBioMedicine, Vol. 67, https://doi.org/10.1016/j.ebiom.2021/03358.

1007/s00521-019-04 https://doi.org/10.1016/j.cobeha.2017.11.

27. Pardos, Z. A. et al (2019). "Data-Assistive Course-to-Course Articulation Using Machine Translation," In: Mitchell, J. C. & Porayka-Pomsta, K. (eds.) In Proceedings of the 6th ACM Conference Learning.

28. Piety, P. J. et al (2014). Educational Data Sciences: Framing Emergent Practices for Analytics of Learning, Organizations and Systems: Race (Conference Session). Fourth International Conference on Learning Analytics and Knowledge, Indianapolis, IN.

29. Prates, M. O. et al (2020). "Assessing Gender Bias in Machine Translation: A Case Study with Google Translate," Neural Computing Applications, Vol. 32, No. 10, pp. 6363-6381, https://doi.org/10.1007/s00521-019-04144-8

30. Reardon, S. F. (2016). School District Socioeconomic Status: Race and Academic Achievements, Stanford Center for Policy Analysis

31. Reardon, S. F. & Stuart, E. A. (2019). "Educational Research in a New Data Environment," Journal of Research on Educational Effectiveness, Vol. 12, No. 4, pp.567-569.

32. Rosenberg, J. M. et al (2020). "Making Data Science Count in and for Education," In: Romero-Hall, E. (ed.). Research Methods in Learning Design and Technology, Routledge, pp. 94-110.

33. Salganik, M. J. et al (2020). "Measuring the Predictability of Live Outcomes with a Scientific